Us, AUGMENTED

IMPROVE HOW WE WORK, ORGANIZE, AND INNOVATE BY AUGMENTING INTELLIGENCE, INDIVIDUAL AND COLLECTIVE, WITH THE NEW AI

WHITE PAPER

GIANNI GIACOMELLI

SUPERMIND.DESIGN

Version 2.0

Last edit: June 14, 2025

First published: December 2, 2024

Download the latest at **www.supermind.design**/resources

Access the digital twin on OpenAI's Custom GPT store **here**.

Contact: info@supermind.design
https://www.linkedin.com/in/ggiacomelli/

About Supermind Design

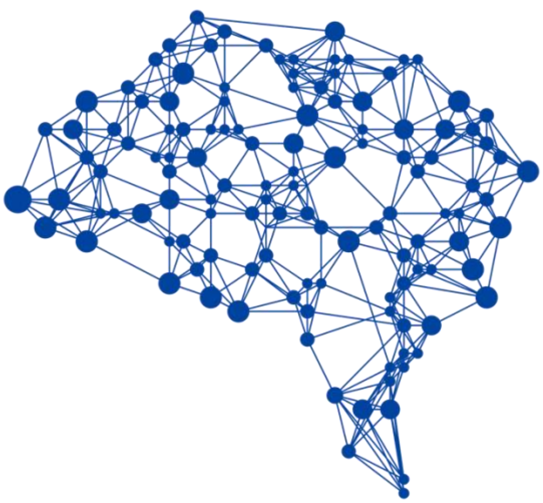
Contact us if you want to use AI-Augmented Collective Intelligence to design and build the next generation of what your organization and teams do.

At supermind.design, you will find resources, from workbooks and videos to a database and interactive AI tools.

Let's go build superminds.

Table of Contents

Summary and Introduction



Preface

In times of fast change, it is important to embrace frameworks that help us make sense of what's happening. This paper uses the concept of "superminds[1]," AI-augmented Collective Intelligence (ACI) systems that make us effective professionals, organizations, and societies: they create new and improve existing ideas at scale and speed; they help the world know what the world knows and learn from what the world learns; they offer a here-and-now counterpoint to Artificial General Intelligence (AGI).

Seeing the world from that standpoint, there is a possible future where AI gives us superpowers not just individually but, even more importantly, it augments us as teams, groups, organizations, communities, and ecosystems. A future where we manage to mitigate and control the risks AI poses to our society and economies. We can design and innovate into that future. This paper is about the journey's why, what, and, to some extent, how – especially when designing the interplay between people and AI machines at scale.

Using this lens, some significant insights have emerged. You will find them described in the paper's articles, organized in three main sections:

- How we organize
- How we innovate
- What bigger-picture impacts are likely.

These find practical application in designing organizational processes and structures, and evolving skills, knowledge management, and collaboration infrastructures. These are "high-leverage points" in the system dynamics of companies, organizations, and ecosystems.

We can design and build them.

The summaries of the corresponding chapters are in the next pages, to help you navigate through the document.

This is a living document, and its current version, as well as its companion interactive tools, are always available at www.supermind.design/resources as well as through this paper's *digital twin* (see the respective chapter.)

Enjoy the read.

Gianni Giacomelli

[1] Prof. Thomas Malone, Massachusetts Institute of Technology, Center for Collective Intelligence, and his book "Superminds"

About This Paper

This is a collection of insights and provides practical ideas for transformation. It builds upon recent advances in AI, the research on AI and collective intelligence done at MIT's Center for Collective Intelligence (CCI) Design Lab, and the extensive "sensemaking" performed elsewhere in academia and industry.

It focuses on the impact of AI, but attempts to provide insights that last much longer than the even-shorter headlines and technology-releases cycles – insights on how people, processes, and organizations will need to respond to the current exponential change.

This is *not* an academic paper. It explores these themes using practical experience from decades of designing and building innovation at the intersection of technology and people's capabilities in large and complex organizations.

Us, Augmented is for a general management audience (executives, leaders, managers, employees) and first sketches out the impact AI-Augmented Collective Intelligence (ACI) could have on our organizations and ourselves by summarizing dozens of stories from 2030—narratives of innovations that improve our ever-challenging world. The paper then discusses how ACI impacts crucial management topics, from skills and training to knowledge management, collaboration, and organizational (especially process and related service and product) design.

Ultimately, "Us, Augmented" aims to move the discourse and practice beyond starry-eyed tech utopia and rampant dystopia that stoke anxiety and promote helplessness.

The paper also reflects the realization that we don't have an option not to play the AI game. Whether because some of our strongest competitors will be organizations that effectively embrace AI or because our society and economies have already accumulated debts that require an acceleration of innovation, there are reasons to focus on AI for the future of work and the transformation of our organizations and processes.

What if we got this future right? And what does it take to make that happen?

Chapters Summaries

ENVISIONING A FUTURE WE WANT

"**Future Back: What Could Happen If We Get It Right?**" presents a positive outlook on how Augmented Collective Intelligence (ACI) could improve the world by 2030. ACI would combine human and AI capabilities in intelligent networks, forming "superminds" that could tackle complex problems in areas like work and education.

IMPROVING PEOPLE, WORK, ORGANIZATIONS – AND AI – WITH AUGMENTED COLLECTIVE INTELLIGENCE

Four Effects Drive the AI-powered Future of Work — machine capability, process redesign, scaled output, novel pursuits: (1) machines will keep taking on more tasks; (2) processes will be rebuilt around them; (3) today's work will explode in volume and reach; and (4) brand-new activities will appear. This shift will push organizations toward fluid, market-like networks where people specialize in problem-finding and purpose, while machines handle how.

Generative AI's #1 job: worker augmentation: Generative AI can augment workers by improving their skills in various domains, regardless of their current level of expertise. It can make laypeople smarter and help experts reach new heights – but the ways it does so aren't the same.

Can AI Make Us Great Beginners at Everything? AI can quickly provide basic knowledge in new fields, making people better beginners and more versatile. This enables individuals to contribute to a wider range of tasks, leading to more efficient and especially effective organizations.

What to Learn in the Age of AI: In the age of AI, "augmented thinking" is crucial. This skill set emphasizes critical thinking, problem-solving, creativity, collaborating with AI, and understanding collective intelligence.

Three Skills Get GenAI To Do More For You: To effectively leverage GenAI, we must master keeping humans (and ourselves) engaged, guiding the machine through a deliberate thought process, and using effective prompting techniques.

Superhuman Knowledge Workers? AI Exoskeletons and Scaffoldings. To control, harness, and compete with AI (and AI-powered organizations), we can make knowledge workers more effective, efficient, and more satisfied - individually and in teams through "scaffolding" AI, which supports skill-building and gradually reduces assistance, and "exoskeletons," which provide continuous enhancement.

Capability + Effort: Exploring AI's Jagged Frontier: Identifying AI's "strike zone" depends on more than just comparing human vs machines' capability. It also depends on the amount of capacity (effort) expendable. A new to scoring individual activities' readiness.

On AI's Dislocation Of Human Labor, We Owe The Workforce Better Guidance: Organizations need to guide workers through AI's impact on jobs by identifying AI-affected tasks, providing training for necessary future skills, and adapting HR practices to manage AI-augmented work. Failure to do so will result in workers' opposition to innovation activities, among others.

With AI, Learning and Reskilling ≠ Training: AI necessitates a shift from traditional top-down training to continuous learning in the flow of work through peer networks, requiring organizations to identify and provide relevant skills for human-AI collaboration, leverage AI-powered tools for knowledge access, and adapt learning processes.

Beyond "Human in the Loop": Reliable AI in Enterprise Workflows: To improve AI's reliability in enterprise settings, organizations should adapt established process management practices, such as task analysis, quality control, and performance monitoring, for managing and supervising AI-augmented workflows. Inspired by classic methods whose conceptual relevance has never been stronger, but whose implementation must adapt to the current times.

Leadership in the AI Era — experiment, question, co-decide, architect inimitable systems: Tomorrow's leaders will run rapid AI-powered experiments, frame the incisive questions machines can't, pair human judgment with algorithmic rigor in "cyborg" decisions, and build cultures, learning loops, and network topologies that fuse people and code into advantages rivals can't clone.

Why Some Quit, And Some Stay: A Surprising Take: Employee retention is linked to strong internal networks and opportunities for impactful contributions. Organizations can increase retention by fostering collaboration and providing access to knowledge and tools. The article is particularly useful in times where knowledge access can be overhauled by AI.

Is Your Organization Intelligent?: A classic article with still-relevant principles. Organizations need to adapt to a rapidly changing world by designing their collective intelligence systems. This involves leveraging technology, promoting collaboration, and facilitating knowledge sharing to enhance organizational learning and decision-making.

AI's Human Side: An overview of what "collective intelligence" means in the age of AI – written before the Generative AI era, and still relevant. While AI plays a crucial role, the success of many tech-native companies also stems from their harnessing users' collective intelligence. Organizations can leverage this insight by fostering collaboration, knowledge sharing, and user-driven innovation.

Lessons from the Past: How Augmentation is Hindered. Enterprise-AI déjà vu — immature MLOps, dirty data, legacy stacks, linear innovation, siloed talent, lax governance: Many projects in the 2010s stalled at pilot-purgatory, but the few that fixed six gaps rewired their cores and captured outsized value—use those lessons now so GenAI and agentic systems scale instead of sputter.

IMPROVING INNOVATION AND PROBLEM SOLVING – WITH AUGMENTED COLLECTIVE INTELLIGENCE

Problem-finding AI agents and Exponential Serendipity — idea-structure mapping, collision engines, knowledge-halo scouting: Autonomous “problem scouts” parse the why/what of ideas, smash them against external challenges, trade insights with other agents, and surface net-new, solvable opportunities—turning serendipity into a system and giving organizations early-inflection radar.

GenAI Must Ask Questions, Not Just Give Answers — slow-thinking probes, context-gathering prompts, analogy-stretching lenses, multi-persona critique: Flip the script so models quiz, clarify, and debate before they create—expanding the problem space, countering shallow System-1 replies, and pairing human symbolic judgment with AI’s wide-angle exploration for deeper, bias-checked solutions.

GenAI as a Personal Problem Solver: A Case Study: AI can assist in problem-solving by exploring and structuring the problem space, generating potential solutions, and evaluating their feasibility, illustrated by examples of using AI for business process improvement.

Generative AI Can Ideate Harder: By leveraging human-created frameworks and facilitating the recombination of ideas from diverse fields, AI can generate novel and creative solutions, exceeding human capacity for exploring solution spaces.

Your Problem-Solving Idea Flow, AI-Augmented: AI can be integrated into each stage of the problem-solving process, from understanding the problem's "why" to exploring solutions and refining them, emphasizing human control and active participation in the process.

Ideas “Physics and Chemistry” with GenAI — idea-atom mapping, collision reactors, hardening loops: Treat every concept as a why/what/how molecule, then let GenAI smash those sub-components against other notions, analogies, and critique lenses to spawn derivative ideas, stress-test them, and ignite chain-reaction innovation. Organizations that architect workflows and networks to maximize these AI-catalyzed collisions harvest continuous, scalable breakthroughs instead of one-off sparks.

Humans Fall in Love with Solutions—AI Can Help Fall in Love with Problems — AI-augmented framing, perspective multiplication, bias buffers, structured decomposition, knowledge uplift: Pair human judgment with machine-driven wide-angle scouting to deepen the "first diamond," surface hidden entry points, counter premature convergence, and craft richer problem spaces—so teams tackle the right challenges before solving them brilliantly.

Harden Your Ideas with AI: AI can critique ideas from multiple perspectives using human-made frameworks, leading to more robust and well-developed solutions. This makes the innovation process more efficient by identifying flaws and suggesting improvements.

Relevance is much of what we need from AI - Most of us can solve complex problems because we get the right external input at the right moment, often over long periods — the foundations are in place already, but by 2030, context-aware "whispering" copilots will likely curate real-time smartstreams, filter noise, tag reliability scores, map concept networks in 2-D knowledge graphs, and connect supermind communities—compressing the invention-to-scale cycle.

Better Ideas by Taking Turns with AI — human-first brainwriting, clarifying-question prompts, alternating human/AI inputs, synergy loop: Capture your own thoughts first, invite AI's perspective next, let the model cluster themes, then iterate together so humans stay in control and real creativity beats passive content-watching.

"AI Psychedelics" For Radical Innovation? Like psychedelics, AI can push thinking beyond conventional boundaries. It facilitates access to diverse fields, leading to novel combinations of concepts and radical innovation when guided adequately by human input.

THE BIGGER PICTURE: IMPACT ON SOCIETY

Will AI sharpen or dull our minds? While concerns exist about AI making humans reliant and diminishing critical thinking, the chapter argues that AI can also enhance human intelligence if used correctly. The key, among others, is to design systems that promote active user engagement.

Are we small models? — humans mirror big-to-small model distillation: we're "lightweight" learners fed by the world's supermind, now amplified—and potentially filtered—by AI, promising efficient knowledge transfer yet raising who-tunes-the-filter risks. A very large shift looms, whether we are ready or not.

Us and our machines are lenses – and that matters immensely — dimensionality reduction, lens orchestration, meta-selection, fusion & evolution, human inventiveness, AI recall consistency: Brains and self-supervised models both compress reality into “good-enough” frameworks; next-gen AI can choose, stack, and iterate these lenses at scale while humans craft new ones, knitting a perpetual, pattern-seeking supermind that shrinks uncertainty and unlocks collective intelligence.

We Are GenAI's System 2: AI systems excel at fast, intuitive thinking (somewhat conceptually equivalent to Kahneman’s System 1), while humans can provide deliberate, analytical reasoning (System 2). Integrating both enhances problem-solving and decision-making in strategic planning, complex analysis, etc.

Stop Working Like It's 2019: This article written during the initial work-from-home mandates is still relevant. Organizations need to adapt to the rapidly changing landscape of AI by fostering continuous learning, leveraging new AI-powered tools for collaboration and knowledge management, and rethinking traditional job designs and organizational structures.

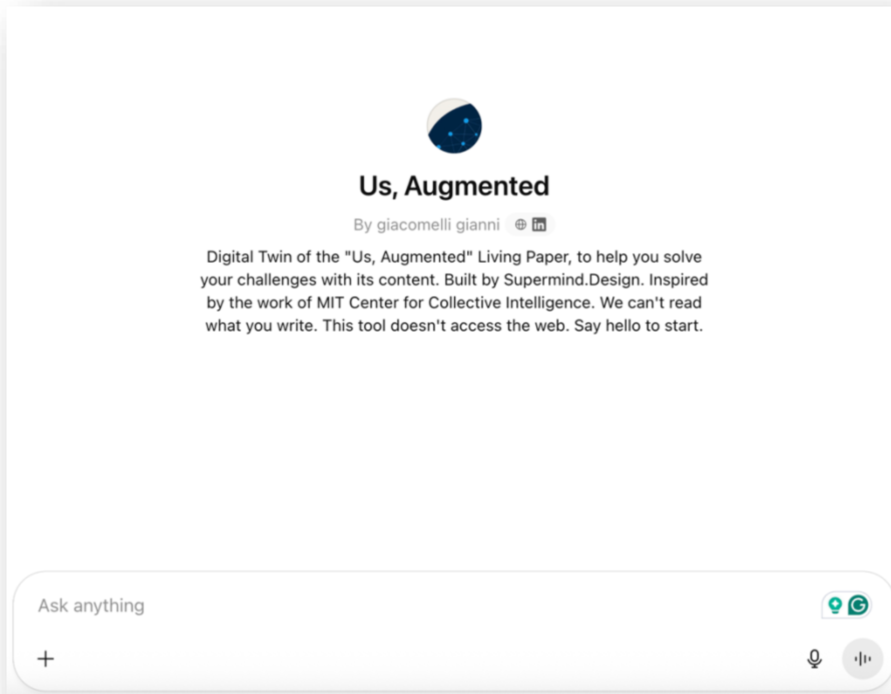
Our Collective Brain Is Ageing. What Does It Mean For Our Civilization?: An aging population presents economic and innovation challenges due to changing cognitive abilities and institutional inertia. Solutions include leveraging AI to augment worker productivity, promote intergenerational collaboration, and adapt governance structures.

Cut Climate Tech Invention-to-Innovation Time: To accelerate the adoption of climate solutions, specialized knowledge-sharing platforms are needed. These platforms can leverage AI and collective intelligence to connect experts, practitioners, and relevant information.

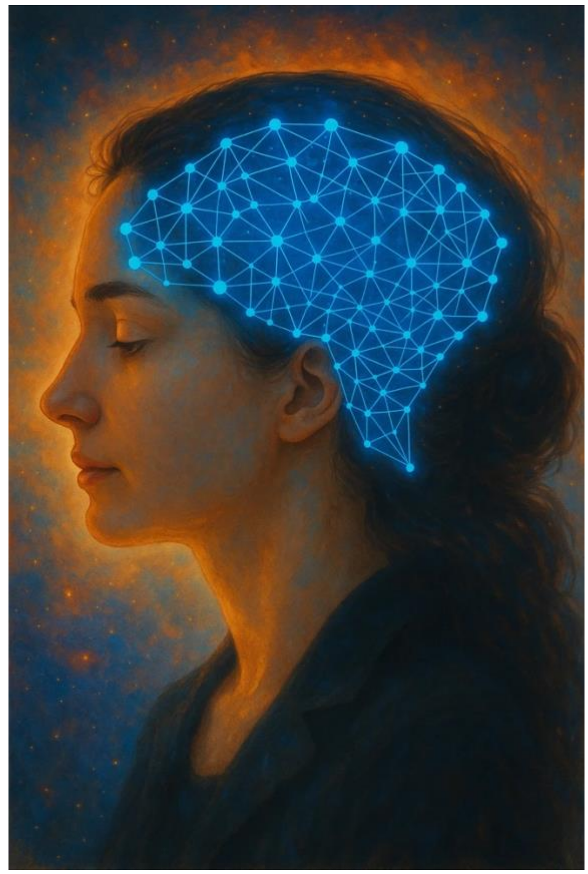
If The World Knew What The World Knows: Humanity's collective intelligence is hindered by the limitations of knowledge sharing and organizational structures. Investing in AI-powered "supermind utilities" can improve access to knowledge and foster innovation. 🧠

Digital Twin of the paper

Available on the OpenAI Custom GPT Store
as “Us, Augmented” [here](#).



Why Augmented (Individual and Collective) Intelligence - Envisioning a Future We Want



Future Back: What Could Happen If We Get It Right?

Imagine 2030. For once, no dystopian future. Below are short fictional vignettes from a time that has harnessed and augmented the power of collective intelligence by tapping into the wisdom and creativity of networks of humans – people, us – supported by intelligent machines. They are intended to guide our innovation design and uncover meaningful opportunities for doing things we don't do today.

In that future, despite the significant challenges, we have understood how to sustainably improve critical parts of our lives in these possible futures. To achieve this, we turned some of our planet into a superbrain—literally—and addressed our complex problems with that intelligence.

These ideas' objective is to *inspire* innovators – products and services, private and public. They aren't intended to be perfect from a desirability, feasibility, or viability perspective. That would have required much more effort and time. As with all powerful technology and innovation, much could go wrong – from human rights abuse to manipulation.

However, *your* expertise can complement Augmented Collective Intelligence (ACI) and its design principles.

Just imagine, for a moment, that we get the "how" right – what could we do with it?

Intelligent networks are made of many people and AI-powered machines connected in a distributed architecture. That's ACI. From Wikipedia to Reddit and YouTube, from Patients Like Me to Bitcoin and Apple's technical communities, and from Pinduoduo to Haier and Bellingcat, hundreds of intelligent networks help harness the full collective cognitive power of people, organizations, and ecosystems. They complement or substitute traditional, hierarchical structures.

These are what some call "superminds[2]". New organizational designs can capture their emergent intelligence. Building one means enabling a network – not one person or machine, or a few – to sense, remember, create, decide, act, and learn. It means helping nodes connect, incentivizing them, and supporting them with knowledge and collaboration tools.

AI can support all of that, by helping nodes to discover each other and connect, curating knowledge, and performing any other computation required. Here are some examples of what ACI superminds could do in 2030 – more in the Positive Futures 2030[3], first published at the end of 2022 and looking increasingly plausible. Here is a list of some of the themes and ideas this

[2] First envisioned by MIT's Prof. Thomas Malone

[3] www.supermind.design/resources

perspective and architecture can make possible:

Transforming the Workplace: ACI is reshaping work dynamics, fostering collaboration, and enhancing productivity.

- **Smartstreams** efficiently route questions and information to the right people, minimizing noise and wasted effort.
- **AI-powered wearables** could monitor user engagement and prevent technology addiction.
- **Idea Colliders** blend ideas from different fields, guided by human-generated and AI-generated frameworks, revolutionizing brainstorming and leading to quicker breakthroughs.
- **Enhanced Collaboration Tools** facilitate seamless co-editing of documents and software across organizational boundaries.
- **AI-Powered Job Matching and Training** connects individuals with suitable jobs and provides access to personalized training materials from diverse sources.8
- **"Professional Identity" Bots** could facilitate networking at conferences and events, promoting meaningful interactions between individuals with shared interests.

Reimagining Human Connection: ACI enables new forms of communication and strengthens existing social bonds.

- **3D Emojis** could offer more expressive and personalized ways to communicate emotions.
- **Enhanced Life Stories:** AI assistants could weave together memories from personal archives and social networks, creating immersive experiences for individuals and families to relive their past.
- **More-Human Local Communities:** Flexible work arrangements and virtual communication tools enable people to stay connected with family and friends while living in different locations.

AI-Powered Knowledge and Learning: AI transforms how people access and interact with knowledge.

- **Search 2.0:** Search engines use AI to surface more accurate and insightful results, including summaries, highlights, and visualizations.
- **Living Books:** Books are no longer linear but interactive, offering multiple paths for exploration based on reader preference. Users can also pose questions to the book, or even chat with the author.
- **Personalized Language Tutors:** AI-powered tutors provide

customized language learning experiences.

- **Social Media 2030:** Social media evolves from a chaotic and often harmful platform to a more curated and intelligent experience.
- **Smartstreams:** Filter out noise, connecting users with people in their areas of interest.
- **AIxtensions (AIX):** AI-powered tools could enable users to engage with climate change deniers and trolls using fact-based and empathetic arguments.
- **Enhanced Reality Experiences:** Augmented and mixed reality technologies become increasingly pervasive, creating immersive and engaging experiences.
- **Realistic Virtual Travel:** Users can experience travel virtually, including realistic 3D environments and even share those experiences with others.
- **Augmented Reality Overlays:** Superimpose realistic visualizations of old buildings onto the real world, allowing users to experience history more tangibly. AR layers could also be used to enhance mental health or make transportation more seamless.
- **Community-Made Art and Entertainment:** ACI empowers creative communities to produce high-quality art and entertainment without large budgets.

Harnessing Trans-species Intelligence: Humans are beginning to tap into the collective intelligence of other species to enhance their understanding of the world.

- **Trans-species Perception:** Augmented reality tools could allow humans to experience the world from the perspective of other species, seeing what they see or hearing what they hear.
- **Bio-inspired Innovation:** The "natural experiments" of biological ecosystems inspire innovations in areas like manufacturing and materials science.

Augmenting Human Capabilities: ACI extends human abilities and senses, creating new possibilities for experience and understanding.

- **Swarmsight** aggregates video feeds from multiple individuals, enabling enhanced situational awareness and real-time collaboration in complex environments like firefighting and construction.
- **Haptic Technology,** while still in its early stages, could allow for a limited sense of touch at a distance, with potential applications for connecting loved ones or experiencing virtual environments.

Addressing Societal Challenges: Collective intelligence is used to address challenges such as climate change, misinformation, and social injustice.

- **Climate Change Superminds:** Networks of experts and citizens collaborate to find climate change solutions, using AI for knowledge management and engagement.
- **Combatting Misinformation:** AI-powered tools and crowdsourced initiatives could help identify and counter the spread of misinformation.
- **Broad-Based Lawmaking:** Democratic legislation benefits from input from expert and citizen communities, ensuring more informed and representative policymaking.

Reinventing Education: ACI transforms education by making learning more accessible, personalized, and relevant.

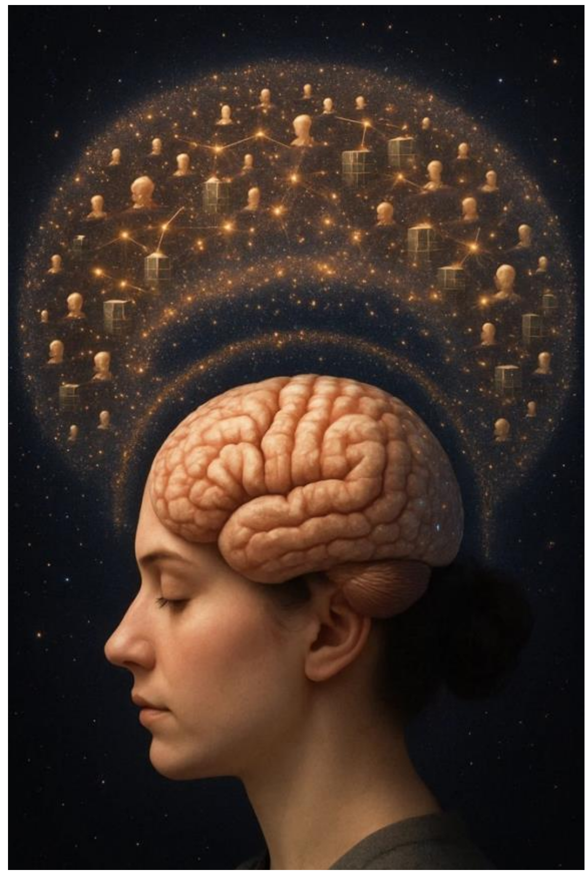
- **Global Research Marketplaces** could connect students with meaningful research projects beyond their local institutions.
- **AI-Powered Curriculum Development** identifies knowledge gaps and rapidly integrates new information, aligning education with society's evolving needs.
- **"Wikipedia 2030" Learning Model** could offer a collaborative and constantly updated approach to education, accessible to a broader audience.

Promoting Transparency and Accountability: ACI is leveraged to build trust and counter negative societal trends.

- **"Halls of Shame"** could use AI and crowdsourced information to document and expose individuals or groups who have engaged in harmful activities, serving as a deterrent.
- **Public Figure Stance Assessment** could employ AI-powered knowledge graphs to analyze public information, offering insights into the views and actions of politicians and other influential figures.
- **Deepfake Detection** could use AI to identify and flag AI-generated deepfakes, mitigating their potential to spread misinformation and harm individuals.

For the full text of these stories, Read Positive Futures 2030 at www.supermind.design/resources

Improving People, Work, Organizations – and AI – with Augmented *Collective* Intelligence



Four effects drive the AI-powered future of work and organizational design

I believe there are four main vectors of how AI will change how we work, our processes, and our own usefulness as professionals. We can plan around them.

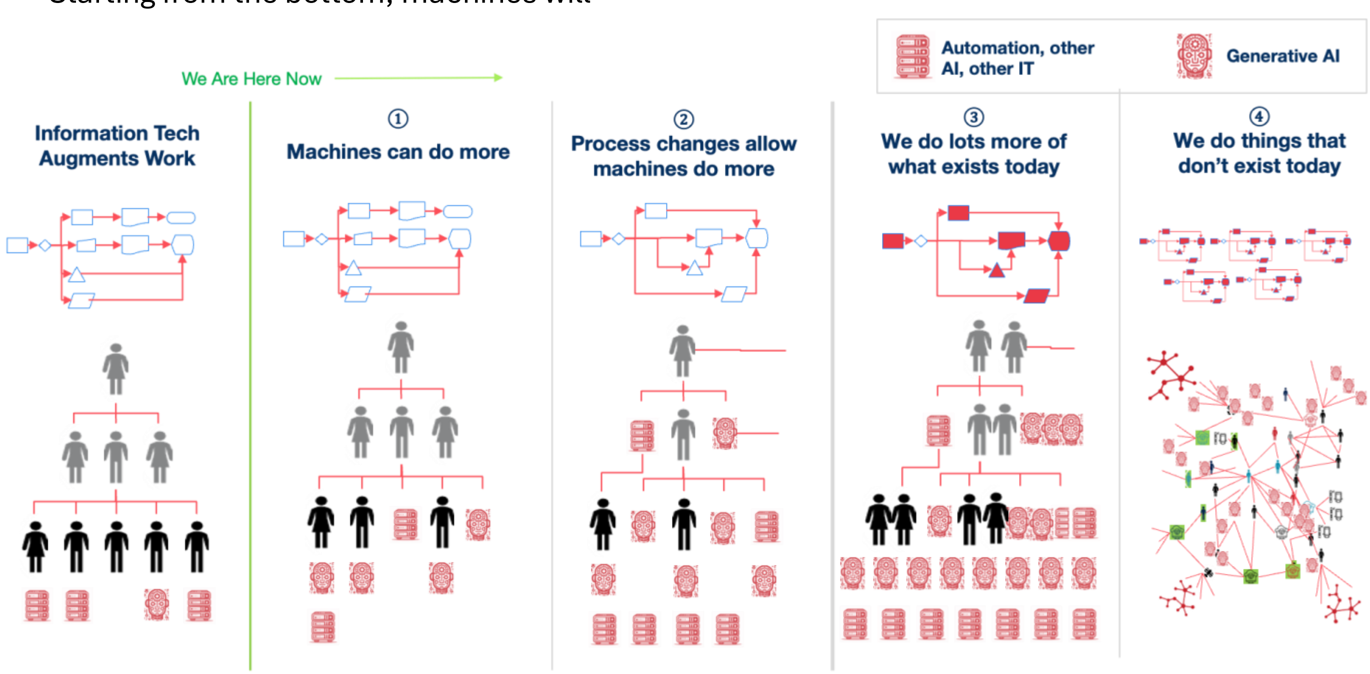
Today, information technology primarily augments work through varying degrees of automation. This includes workflows, generative AI, and predominantly predictive AI from the previous generation, which is still being embedded and developed. Many are still figuring out how to use these machines.

The future:

Driver 1: Machines Will Do More: Starting from the bottom, machines will take over tasks that, within current processes, are increasingly within their capabilities. These tasks will continue to expand as machines grow more capable.

Driver 2: Process Changes Enable Machines Do More: This means progress will not rely solely on machines becoming smarter but also on designing processes that better suit the capabilities of today's machines. For instance, new human-in-the-loop or machine-in-the-loop systems for quality control.

Driver 3: We Will Do More of What We Do Today: We will see a significant increase in what we already do today. Current processes and services will become more affordable, enabling greater scale. For example, we'll produce much more software code and content—text, multimedia, and beyond—which will improve overall output and productivity.



Driver 4. We Will Do Things We Don't Do Yet...provided we have the necessary resources, such as energy. Looking back 100 years, many things we now do and desire were unimaginable. Similarly, we should expect new, unforeseen pursuits to emerge, and scaled supply follows.

As a result, I anticipate a profound shift in organizations and people. Beyond the increased presence of machines working alongside humans, the sheer volume of work will also require more—and new—human involvement. For instance, while machines will write more code, we will still need developers to identify the right problems to solve and quality control.

We will likely see much more dynamic, new networked processes and organizational structures. These structures will allow agents to connect in ways that redesign processes more fluidly than current management practices permit. Some of these organizational forms may resemble markets, where demand and supply are balanced at a granular level. In such scenarios, agents—whether human or machine—could leverage resources to achieve increasingly diverse objectives.

Despite these possibilities, the exact role of humans remains unclear. My sense is that people will focus more on the "why" and the "what," becoming problem-seekers and helping machines scope problems effectively rather than serving as the executors of the "how."

Many of us are unprepared for this shift; our current training infrastructures cannot support it - and our organizations don't learn enough. I have written about it extensively elsewhere.

The time to start preparing is now.

Generative AI's #1 Job: Worker Augmentation

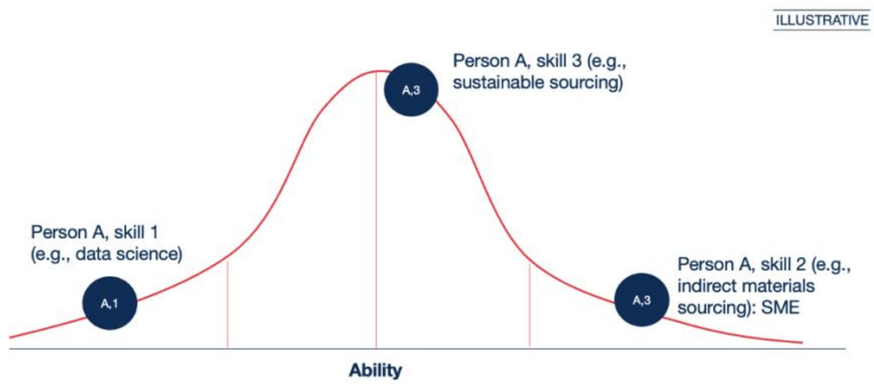
There is still too much focus on Generative AI's ability, or lack thereof, as a valuable, reliable standalone *automation* tool. While those capabilities are important and will continue to improve, they shouldn't be the center of today's attention.

Instead, we must focus more of our discussion on using generative AI to *augment our capabilities* as *individuals* and *groups* because that works well already. While there are overlaps between them, capturing AI's power for individuals and groups presents differences that we must account for in our organizational and process design and our guidance to employees (for instance, as part of their learning curve).

I wrote elsewhere about process design to harness Generative AI and augment organizational capabilities. This article will focus on *people* augmentation and make some simple observations that should help leaders - organizational and process designers and individual professionals - identify the opportunity and pursue it practically by choosing the right AI-augmented tools and practices.

Augmenting individual capabilities

Everyone's capabilities on anything vary from the relatively least to the most developed. Compared to others in a specific professional domain, those capabilities are likely distributed on some curve – let's assume a bell-shaped one, for simplicity. See the illustrative example below of a sourcing subject matter expert, who is in the top quartile for, say, indirect material sourcing but in the middle of the distribution for sustainable sourcing practices, and at the bottom for data science.

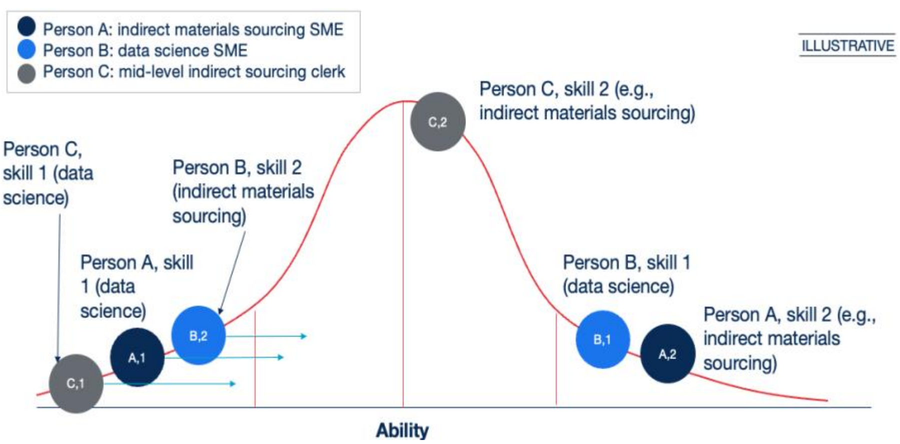


That person's work, done individually, would likely benefit from improving their sustainable sourcing capabilities and their data science practices. This is the focus of many traditional learning infrastructures, with their courses and instructor-led teaching. However, as I argued elsewhere, that is not sufficient anymore in a world where skills change fast, especially as AI shifts the boundary between human and machine's work.

Augmenting collective capabilities

While augmenting individual capabilities is a problem worth exploring (and we will in a moment), it is not the most significant challenge. As we all know, **much work happens in groups**, and that's where we need to help most people.

The visual below illustrates a typical situation in which a hypothetical group of three people (a data scientist, an indirect materials sourcing clerk, and a sourcing subject matter expert), with their respective skills and gaps, join forces to solve complex problems that elude individuals' capabilities in isolation. Their weakest individual capabilities are augmented. This could happen in an escalation, continuous improvement, or transformation project.



The upshot is a *combination* of the collective capabilities of the group, not only *offsetting* each other's weaknesses but also supporting the constructive debate and ideation that leads to making everyone – even the most competent in the team – smarter. This is crucial for complex problem-solving. In this type of scenario, the least skilled people in any specific area must come up to speed quickly, to interoperate effectively with other professionals. In this example, a better understanding of indirect material sourcing benefits Data Science subject matter experts: it helps them collaborate more effectively with both business domain (sourcing) subject matter experts and the clerks who will ultimately use the new processes and systems most. In essence, *anyone's least-developed* (yet relevant) *capabilities are the constraining factor* in collective problem-solving.

How does Generative AI help in each of these scenarios?

In a nutshell, AI, especially the generative AI type, can help in two main ways:

(1) making people in the bottom quartiles smarter so they can do what was done by mid-distribution professionals and

(2) allowing those in the top quartile to go beyond what humans (*any* human) could do before.

That also likely means that generative AI will relatively organically boost the performance of companies least endowed with certain capabilities (say, marketing). The performance of those at the top could benefit, too, if their top experts cannot only improve their least-developed skills but also embrace new ways of working that harness the harvesting of cross-disciplinary ideas.

We will review some examples in more detail below. First, though, we will break the problem into different types, as there are subtle differences depending on the type of work: "run" (daily operational work), "build" (improvement and transformation delivery), and "design" (coming up with new ideas to inform that transformation delivery). Let us explore them one at a time.

In the flow of daily operational work

"Run" work is typically quite streamlined and requires group interactions, mostly for exception management and escalations. For example, consider the need for better disclosure of carbon emissions by specific suppliers if they are not already prequalified and their reporting is unconvincing. In those situations, mid-level employees may need to interrogate analytical tools to estimate the actual risk and/or reach out to subject matter experts to validate those risk assessments and agree on the next step.

In these circumstances, a certain amount of domain expertise *across* all people involved is essential for an efficient issue resolution across all people involved. Employee attrition and loss of tacit knowledge are typical challenges. To this end, the typical process design includes giving access to basic training and self-help tools, leveraging knowledge management, and then streamlining synchronous and asynchronous interaction between supervisors and subject matter experts with the help of collaboration technologies, such as Microsoft Teams, Slack, Basecamp, etc.

In this scenario, generative AI has clear opportunities to support human groups. Consider the following, among others:

	Brief Description	Examples of Application
Facilitate Access to Knowledge	Generative AI aids in learning and knowledge retrieval at point of need.	Using AI like Perplexity.ai for engaging learners, suggesting analytical steps in Excel, enhancing understanding of complex documents, transcending language barriers.
Bolster Collaboration	AI enhances collaboration by documenting and organizing communications and assisting in various interpersonal tasks.	Memorializing meetings, drafting multimodal communications, assisting in negotiations and feedback, organizing events, acting as a therapist or entertainer, and helping overcome language barriers in asynchronous communications.
Critique Reasoning and Decisions	AI supports decision-making by presenting options and detecting logical gaps.	Troubleshooting issues with pros and cons, using a Socratic method to improve decision quality, detecting blind spots in logic, potentially leading to automation of tasks and reshaping job roles by augmenting less expert co-workers and unbundling skills of subject matter experts.

Facilitate access to existing knowledge. For instance, by supporting the learning of people who are new in their roles. Generative AI is already effective at engaging learners by retrieving contextual knowledge (think of Perplexity.ai as an example) or probing learners through questions and quizzes. Generative AI combined with better search capabilities, as well as knowledge graphs (which identify both content-type *content* as well as *people* with relevant skills), will likely become a great fact-checker and even more importantly, a curator of relevant knowledge that can be presented to humans at the point of need, and that can be *engaged with* – not just read. Generative AI can already suggest analytical steps, such as formulas in Excel sheets, to dig deeper into data - and possibly help write basic code for more complex data science work that laypeople could now perform. Similarly, it can help laypeople quickly improve their understanding of complex legal documents or even inspect the behavior of software code. It also helps people transcend some of the language barriers. Also "one person's automation is another's augmentation": think of nurses who, by using tools powered by Generative AI enabling better diagnosis and prognosis, as well as therapy adjustments, can start doing some of the work previously done by doctors; and in the same vein, younger and less experienced nurses could mitigate some of the bottlenecks due to the scarcity of experienced ones.

Bolster collaboration. For instance, by practically memorializing and summarizing conversations, including those that happen on audio and video, cataloging them to make them easier to retrieve for those who participate in those meetings, and simplifying the drafting of (multimodal, including images, video, audio, and web microsites) communications. Also think of AI's possible role as a coach to nudge people to the right behaviors in interactions, a mediator in disputes, an assistant in negotiations, a sparring partner for a dry-run before a difficult meeting, a helper in providing performance feedback to colleagues, an organizer for retreats and workouts, and eventually even a readily available occupational therapist, or an entertainer to keep the atmosphere lighthearted. As in the previous point, Generative AI also helps people transcend some of the language barriers, especially in asynchronous communications.

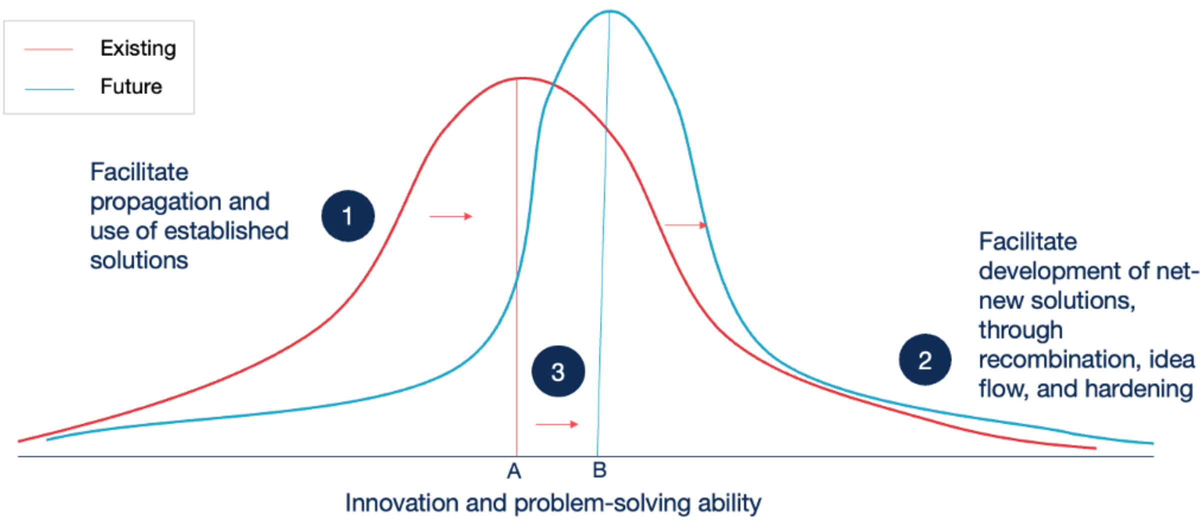
Critique reasoning and decisions. While Generative AI is often not accurate enough to make decisions independently, it could be used to present possible options and troubleshoot issues, with a list of pros and cons, to human deciders. It can help detect blind spots in human logic or support peer software development code reviews which are often neglected. Even more intriguingly, it can be used to ask questions to humans, playing a Socratic dialectic that helps improve the quality of the decisions and makes employees learn. Over time, these capabilities may become good enough to lead to actual automation of tasks, which in turn might reshape job shapes (for instance, by helping to unbundle scarce and expensive skills - like part of those of a subject matter expert - and augment the work of less expert co-workers).

All of the above may or may not be confined within the boundaries of an individual organization, which makes the effective processing of information and the ability to collaborate even more important.

As part of innovation and transformation work

“Design” and “transform” work typically either improves processes based on some type of continuous improvement or reimagines them. In both cases, deeply understanding and empathizing with the experience of the stakeholders involved is a necessary foundation for the effort. In these types of work, Generative AI can help people be more effective and efficient. Consider the following, in addition to the points described above.

Make laypeople a little smarter, fast. Providing people in the bottom and even in the middle quartiles of the capability distribution with a basic understanding of other people’s domains facilitates more productive interactions with experts.



Think of a data scientist with a limited understanding of emissions reporting for indirect sourcing but who needs to productively debate the design of analytical capabilities with both the subject matter experts (who see the trends, the prompt new analytical capabilities) as well as the mid-level professionals who will be using those tools every day. In this example, at the same time, sourcing experts benefit from learning some of the new capabilities afforded by contemporary data science quickly, and from exploring more thoroughly the lived experience of mid-level workers to foresee potential change management issues for instance. That is particularly useful when the work is done with agile methods, because fast ideation, iteration, and identification of requirements is best done when people understand at least some of each other's vernacular. The classic "T-shape" (and generally, any shape but "I") re/upskilling concept was created to facilitate those interactions. Generative AI can help professionals come up to speed to be more effective in interactions with experts, as well as simplify access to technical communications shared by them. Generative AI is quite capable of demystifying technical jargon, for example. An additional, vast opportunity is the use of AI to complement less-developed skills in smaller companies, or those in emergent economies, that lack specialist knowledge (e.g., marketing or HR) whose availability is taken for granted in larger firms.

Facilitate the exploration and combination of ideas from very distant fields. It is well understood that radical breakthroughs typically come from the intelligent and unusual weaving of concepts from different disciplines. Many struggle with that. Generative AI can make them more comfortable and efficient in exploring the frontiers of their knowledge, especially when guided by effective conversation (not just prompt) engineering protocols (see our MIT CCI Ideator work), and knowledge graphs.

Generative AI is also already quite effective in breaking down problems into their components, providing analogies, and examining challenges through human-created lenses such as management theories. For more examples, see our MIT Ideator (registration required) and Apta GPT (OpenAI subscription required).

Strengthen third-party input critique. All of this means not only increased idea flow partially by more effective recombination of varied ideas. It also means hardening those ideas at every step of the process: grounding them in a more thorough collection and exploration of user feedback, exposing them earlier to stakeholders' perspectives in the form of synthetic user personas simulating the ideas' reception by those stakeholders, and reinforcing them by addressing their shortcomings through iterations with synthetic and real users.

The compounding impact on an organization's innovation ability

The compounding effect of all the above is summarized in the next chart. The cumulated capability stock of a company and its ecosystem increases with two different motions.

First, established solutions are propagated more effectively and shift to the right the part of the distribution curve that sits in the bottom quartiles. This effect may be already underway in many developing economies' organizations and workforces, as they seem to be embracing generative AI faster than others.

Second, those who work at the frontier of the known can now engage more productively with a larger universe of ideas, both from outside the organization and those surfaced by and iterated with their colleagues and broader networks.

Generative AI has a significant role in this, but one that is more subtle than just looking for answers through prompts. The impact will be felt most when companies put AI's power to the service of the individual components of problem-solving and innovation processes – and to the service of those who currently fulfill

those roles, making us smarter
individually and collectively as groups. 🧠

	Brief Description	Examples of Application
Make Laypeople Smarter, Fast	Enhances basic understanding across different domains to facilitate interactions between laypeople and experts.	Data scientists and sourcing experts learning basics of each other's fields to improve collaboration; using agile methods to speed up ideation and requirements identification.
Facilitate the Exploration and Combination of Ideas	Uses AI to blend concepts from varied disciplines for breakthrough innovations.	Guided exploration of knowledge using generative AI, conversation engineering, and knowledge graphs; breaking down problems and providing analogies with tools like MIT CCI Ideator and Apta GPT.
Strengthen Third-Party Input Critique	Increases the effectiveness of idea development and critique through diverse inputs and iterations.	Using synthetic user personas to simulate stakeholder reception, collecting and exploring user feedback extensively, and reinforcing ideas through iterations with both synthetic and real users to refine and validate concepts.

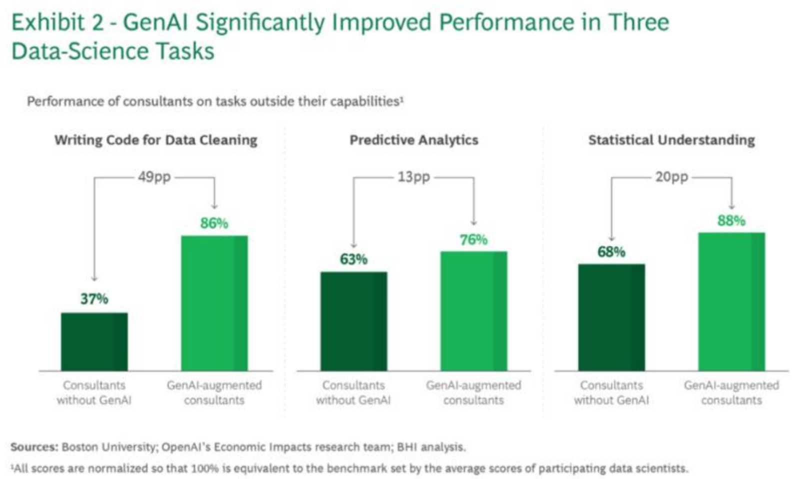
Can AI Make Us Great Beginners at Everything?

Most people, intuitively, think that for AI to be helpful, it should help us improve how we do things *we are good at*. I challenged that assumption in two previous articles (here and here), and it is time to expand on it with additional data.

First, **we do not always perform at our best.** AI could support us for tasks where we have the requisite skills, but for which our attention and effort dwindle because they are repetitive or simply because we may not be willing or able to expend much effort on them (think about reading a 30-page briefing document before a meeting).

Second, **we are *not that good* at many things we do.** AI can augment us when our skill levels are lower. The importance of this point is often overlooked: all of us, in our daily jobs, from frontline managers to CEOs, spend much time on things for which we are mediocre. For many of us, being a better beginner to perform better in areas we haven't mastered is a foundational part of our professional value. There is ground to believe that generative AI can be helpful there. If this is interesting, read on.

A recent study from the Boston Consulting Group empirically showed that professionals **benefit from generative AI when confronted with tasks requiring skills they don't master**. In this specific example, consultants who didn't have data science capabilities could use generative AI to do some data science work.



Source: The Boston Consulting Group

Those who use Gen AI daily probably already know that from direct experience. Some personal, anecdotal examples:

1. Interacting with doctors for benign but long-standing issues that no individual specialist had bothered (or could) look at end-to-end and for which the family doctor needed more data or the time (or the skills) to connect the dots. Think of the interplay between body parts where doctors would deliver a fully effective diagnosis if they were operating as a team, but given their time constraints, they don't. Working with tools like ChatGPT, Claude, Perplexity, Consensus, etc., allows to form additional

hypotheses and helps doctors to generate more holistic treatments

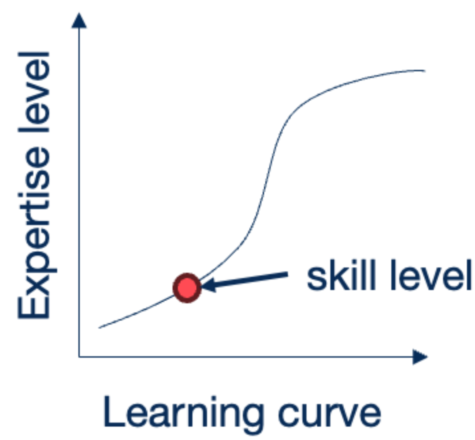
2. Engaging with lawyers and public officials in a new country in their language where you are less fluent. The result is sounding confident and precise, which matters to be taken seriously, understanding the nuances more thoroughly, and generally being able to cover more regulations faster. (Tools used: ChatGPT, Perplexity)
3. Taking cross-disciplinary approaches to research and analysis by connecting dots between themes for which one's knowledge was limited. This applies to both academic research and enterprise innovation, among others. (Tools used: ChatGPT, Claude, Scholarcy, Perplexity, Consensus, Elicit)

Some of these could have been done in the past with "conventional means": search, online translation, etc. However, the speed and breadth offered by the new tools allow us to do much more in the same amount of time, radically changing what one can tackle confidently - and changing how we work.

You are your S-curves skills portfolio

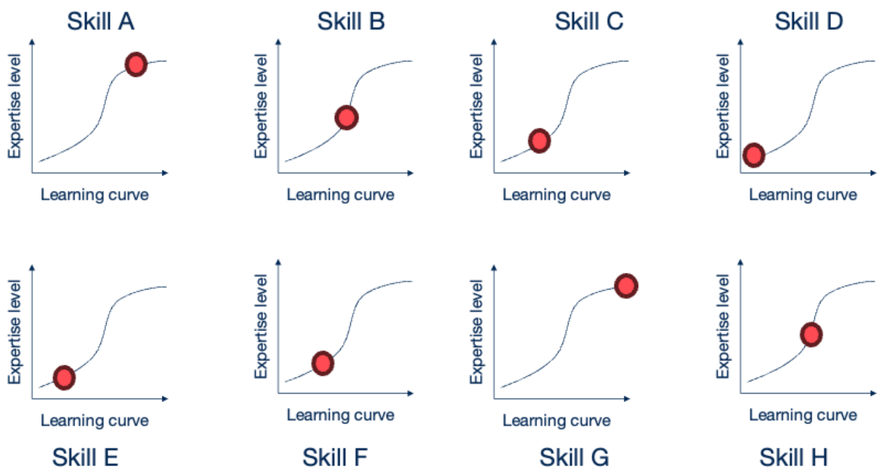
Let's now generalize what this all means.

Our skills and capabilities are crucial to our performance in specific tasks. With effort, strategy, and luck, we reach higher levels in a few chosen skills over time. Typically, only the initial part of skill acquisition is easy (hence the "S-curve" shape); and some skills decay, some are not our forte, and we continue to struggle.



Much of our activities require us to tap into our less-developed skills, like a football player with a less-favored foot (or hand) and field position who needs to operate within an extensive range. In other words, the skills we bring to bear look like a *portfolio* of S-shaped curves, and our effectiveness often depends on their average - and sometimes on the least-developed ones (see chart below).

Depending on who you are or the role you have, you need a specific portfolio:

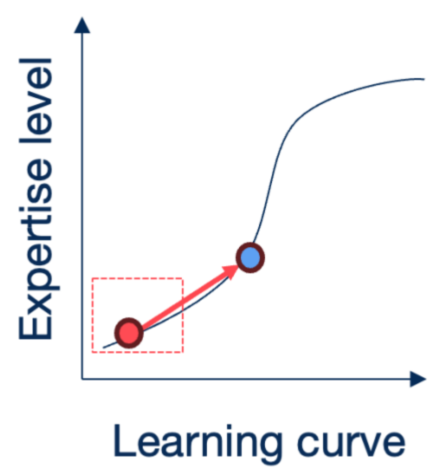


- **Generalists**, like general managers or disruptive-innovation teams, often operate in skill quadrants where they struggle. That's where they delegate or bring teams of experts together, incurring additional "transaction costs" intuiting that investment pays off.
- **Specialists**, like university professors and academic researchers or R&D experts, typically try to spend much of their time in areas where they are world-class. However, they do have to do things that they do less well (say, navigating new types of grant application procedures). Also, in jobs where strong expertise drives most of the success, there might be a need to absorb new knowledge, develop situational awareness, and constantly scan the environment for changes.

Of course, the above is an imprecise categorization, but it gives us an idea of the possible extreme archetypes. Each of us likely falls in a continuum between those at any time.

If used well, generative AI may make us better beginners and help us travel faster and more effectively rightwards from the left to an "advanced beginner" level.

This can have profound implications. Today, it doesn't take us much time to get the basics of new spaces as long as their knowledge exists publicly or within your firm (among others, management consultants have noticed, and many are building out the next generation of knowledge management tools). These new tools also



allow for interplay between us and them, not just to get answers.

If that happens, the portfolio could be significantly improved. Our overall performance is often disproportionally impacted by our weakest spots, and so is our ability to intervene in more areas that require our attention. Learning how to leverage Generative AI to make us dramatically more rounded (in learning lingo, T-shaped, or any other shape than "I") can be a game changer.

The implications of being superhuman beginners

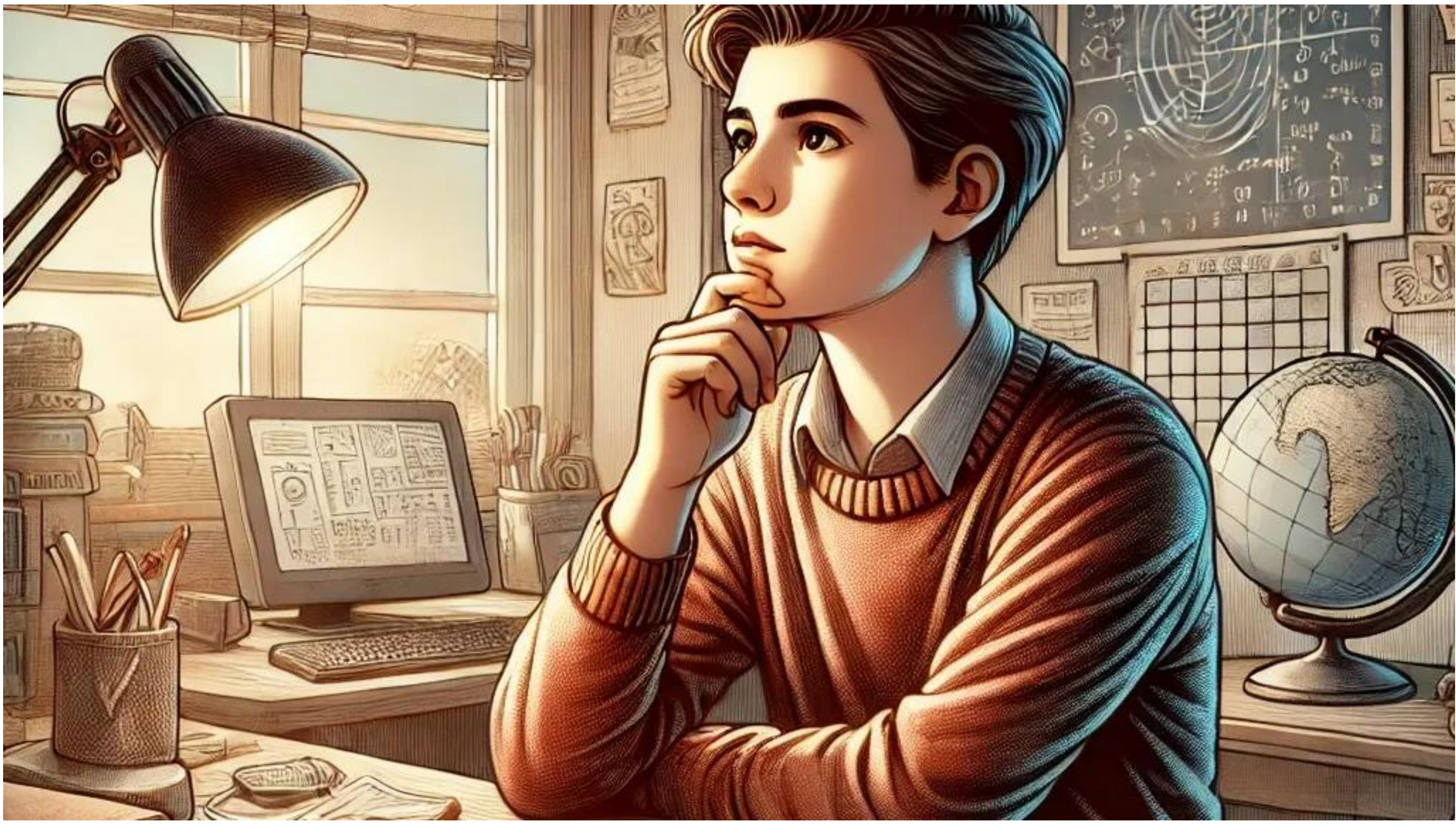
There's a clear "future of work" intuition: traditionally, we were told to stay in our lane, as the transaction costs for acquiring a modicum of new knowledge were high. Experts were tapped into for low-level questions, choking their capacity, and combinatorial innovation (combining ideas from different fields) was very hard. Search engines applied to the internet's fact base changed things. Now, things can change further.

In the longer run, assuming that the tools and the skills to use them improve, this could impact, among others:

- **How we manage:** We can try to understand things outside of our comfort zone without needing to look for experts' help all the time or form better hypotheses to test when experts' limited time becomes available in meetings.
- **Workforce planning and talent management:** We can use more inexperienced or less qualified workforces as long as they learn how to be effectively augmented.
- **How we organize**: making smaller organizations, or ones that are less embedded into traditional knowledge ecosystems, competitive.

More generally, does that mean a possible reduction of the size of effective enterprises (a new twist in Coase's law)? More entrepreneurship? Can developing-countries professionals flatten the first part of the curve more easily? Can young people do more than today? Can combinatorial research and innovation get a boost?

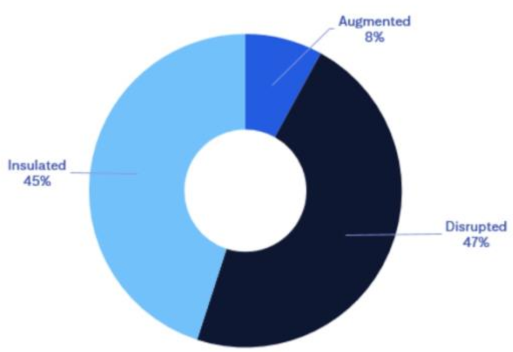
We don't know for sure. Much can go wrong. When using Generative AI poorly, dependency and quality control are risks. However, new "augmented thinking" literacy can help, as can a more innovative approach to organizational, process, and work design. Time will tell, but the opportunity seems large.



What To Learn in The Age Of AI

The capabilities of new technologies improve daily. Our brains and our skills largely *don't*—or at least not yet. And neither do our job designs - leaving many

Figure 6. GAI's expected effect on LinkedIn members' skills, globally



Source: LinkedIn Economic Graph Research Institute

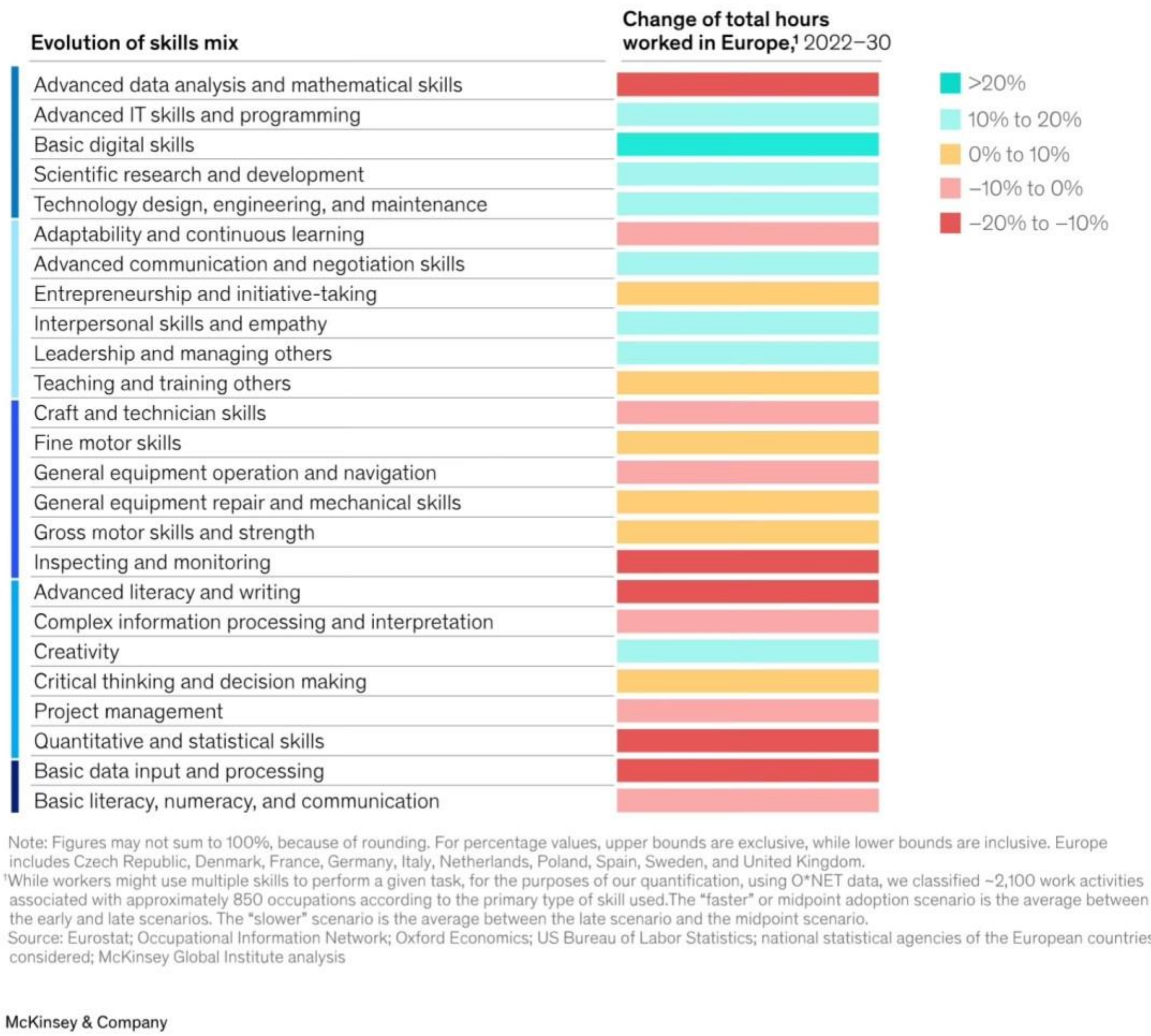
feeling exposed. Charts like the one below referring to Generative AI's impact (and announcements of companies like Klarna repeatedly made) sound everyone's alarm bells.

Via Citi Global Insights

In the "tech + process + people" equation, the *people* side is possibly the weakest right now. That hiatus is causing anxiety, and clear answers aren't always forthcoming.

It is no surprise then that most of us today feel some urge to make ourselves (and our kids) resilient to the change caused by AI, particularly Generative AI's sudden surge. What is, however, particularly problematic is that together with the plethora of claims hinting at doomsday scenarios (and fueling some degree of despondency), we also hear too much wishful thinking and appeasement: "Stay close to what makes us human, things like empathy and creativity." The issue is that the average person's empathy, creativity, or the perception and manifestation of many other cognitive or emotional traits are, on an average day, often no better than today's machines, let alone tomorrow's. That's particularly true when AI is embedded into proper processes designed to automate or augment specific tasks alongside humans. That doesn't mean that all empathy and creativity work can technically and economically be automated. But it means this is no longer a safe, human-only ground.

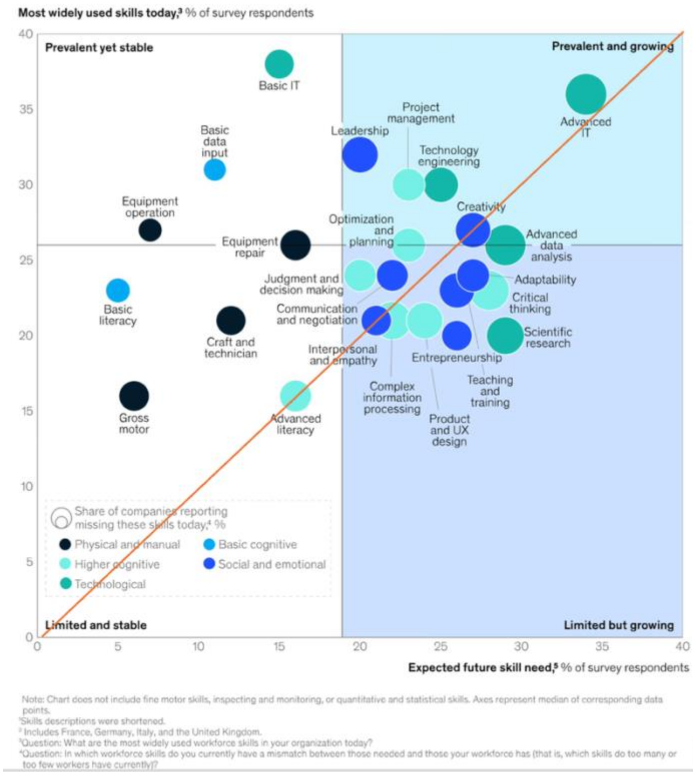
Two datasets from recent McKinsey research (for Europe, but much of it translates elsewhere) are worth looking at. If you believe that you still have many possibly significant career changes in front of you, the first chart is interesting. It shows what type of skills the economy will continue to absorb, given the change in jobs available to people. For instance, data analysis will be easier to do by machines, as will literacy and



communications - and will require fewer hours.

The following analysis is based on a survey. The downside of such an analysis is that today's executives have not fully developed a strong understanding of what AI can do and certainly not what it *will* be able to do. This said, it is useful because leaders and managers focus on existing jobs and may collectively highlight, at least directionally, the incremental steps that workers can take. The skills in the top right box are expected to be in most demand both now and in the next 5 years. The ones below the red diagonal line (which I added) are those that could become more important in the future compared to today.

While there are some differences compared to the previous chart (for instance, data analytics and entrepreneurship), the expected reduction of importance for skills like basic IT, basic data input, equipment operations, basic literacy, and gross motor capabilities is interesting. Similar, though slightly less pronounced, is the dimming of prospects for technician skills, equipment repair, and even general leadership.



Source: McKinsey, 2024

Despite their merits, neither of these analyses gives a full answer, showing the gap in understanding that still exists - a gap that we need to close to provide proper guidance to workers. While a proper answer is being developed and will likely continue evolving, the following principles can guide our individual and collective actions.

Don't try to outrun machines

For historical reasons, much of our training (and, unfortunately, job design) treats us as biological machines. The rote and repetitive learning focused on accumulating notions predates the time when knowledge was easily accessible to those who knew where and how to look. That type of learning counts as little more than mental gymnastics but often doesn't make us smarter than the machines.

Generally, we shouldn't try to outrun the machines on the things they do or will do well. For instance, the remembering of specific notions in isolation. That race can’t be won. That also means many surprising things, including some types of creativity, empathy, etc. For instance, we already know that AI machines are better divergent thinkers than humans in environments where the context is widely available in the training sets of the AI models - for example, coming up with basic consumer product or service ideas. For topics where context is extremely important, such as complicated business strategies or more niche applications, machines will be increasingly fed with the right content as part of creativity workflows, improving their performance.

In numerous areas, AI might become better than most people and, in some areas, all people. What machines lack in terms of representation and understanding of how the world really works, they often compensate through other means, including brute force, especially when paired up with humans who can point them at the right things and help them filter and recombine their own output.

We humans need to get going now.

Today, tomorrow, and after tomorrow

There will be some stability in the short term (less than 3 years) and even longer-term in sectors that continue using legacy processes and systems because they can't change or want to protect workers. That may be unsustainable from an economic standpoint and will also mask the underlying shift, which might prevent those organizations from helping their workers learn. That strategy might lead to a "termination shock" when those organizations can no longer buffer them from the new reality, and the time to adapt will then be too short. In the longer term, say 5-20 years into the future (which is well before when most of us will retire), AI augmentation will be the default choice, and the impact of automation will go deeper when it is within the frontier of the operationally (= technologically + process) possible.

I make no predictions on what happens beyond that horizon. A mere extrapolation of the current trends already takes us into very different territories and misses the likely, sudden impact of breakthroughs—say, the ability to run orders of magnitude more computation because of energy-production discontinuities or computational efficiency.

Understanding what we do before you understand what we need to learn

Let's first try to understand what our work really *is*. A lot of what we do as human professionals can be broken down into three buckets:

Understanding and shaping the "why" of the work. That means forming a clear and actionable view of the reasons why we need to summon organizational resources to do something. That is arguably the job of most senior executives, but it also applies to frontline managers and increasingly decentralized and non-hierarchical environments. Doing a good job at that requires pattern recognition and a continuous sensing of the environment. Formalizing these processes isn't trivial. It requires understanding the interrelations between things in the world and, therefore, a representation of reality that transcends the semantic reasoning that AI models use. It also requires a continuous filtering of irrelevant information. Machines can increasingly complement (see the BCG/HBS research) humans in this process. For example, they can help us evaluate priorities and scenarios and might be able to do some of that autonomously, but it isn't clear how quickly they will be reliable in doing so and at what cost.

Identifying, shaping, and syndicating the "what." This is about matching problems to be solved with the categories of solutions available in the case of known-knowns (defined problems with defined solutions—e.g., how to run a product innovation workshop) or deciding that the problem belongs to the unknown-unknowns category (poorly defined problems for which solutions are not evident or may not even be easy to classify—for instance, how to make people learn for AI's age!). These processes require pattern recognition, including an intuitive understanding of what an organization can tolerate in its change management. Machines can already complement humans here, but humans' ability to think symbolically, with principled representations of the world (e.g., through theories and frameworks), is advantageous.

Identifying, shaping, syndicating, and implementing the "how." Machines are becoming increasingly good and fast at finding solutions for well-defined problems. Here, once more, what they lack in abstracting and using a symbolic representation of reality is compensated by their brute-force ability to connect dots in the semantic space, finding correlations between knowledge that humans have structured in their language. When productively paired with humans, they can help scan a broader horizon of possibilities. This also might mean creating change management plans, where machines can simulate various stakeholders' reactions and help devise personalized change management plans. Or helping humans keep tabs on the change process through more rigorous project management or detecting signals across enterprise communication channels.

So, what should I learn? The rise of "augmented thinking"

In this new world, our role becomes more of an orchestrator, a manager, and a strategist. Much of our work will be on the why and the what, and much of our "how" work will be human-in-the-loop quality control.

The tools will do a lot of the heavy lifting and act as an army of indefatigable interns (and increasingly better at approximating experts' capability levels). That means asking the right questions, including those that lead machines and others to ask you and your networks questions; critiquing questions and answers, individually and collectively; and getting to the right decisions, especially for complex (not necessarily complicated, which machines can tackle more easily) things; seeing patterns and behaviors of systems and using them to guide your and your organization's efforts.

What skills are needed for that? The following is not mutually exclusive or collectively exhaustive, and it is likely no more than directionally correct, but hopefully, it goes well beyond wishful thinking. In the short term:

Critical thinking and its application to framing questions well, critiquing and enhancing answers, general **problem-solving** and creativity, and their AI-augmented use. Humans will still direct many of AI's problem-solving efforts for the foreseeable future. But we need to be good at it and know how to use machines for it if we want the *combination* of us and them to be better than them in isolation

People (and AI labor) leadership and management. This includes the classic and ever more important part of the curriculum - which is also about building good relationships through empathy. But it also includes "network leadership" (the ability to work human networks effectively), modern collaboration tools, human-machine interaction basics, keeping tabs on what machines can do and how to work with them - and helping teams do the same. But beyond the clinical and "economic" view of things, we better heed the importance of mental health and - for leaders - the value of bringing joy into the design and the lived experience of the work

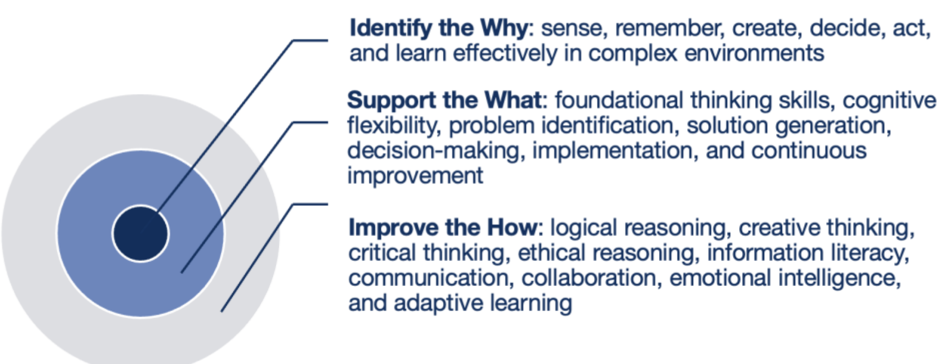
System (not IT systems) **thinking** and the dynamics of large systems that lead to collective intelligence - as that is the architecture of the future. This includes classic skills like social influence but augments them with the requisite understanding of AI-augmented organizational design

Leading and managing the self: Individual resilience, adaptiveness, etc., including adaptiveness in learning, as things will invariably change. Part of this is about metacognition, learning how to learn constantly. Part of it is mental health.

The basics of digital technology. This is not about coding's syntax but rather digital architecture basics and a general understanding of how digital technologies work (including cybersecurity, IoT, algorithmic recommender systems, code architectures, and many others). Those are the building blocks of our future world; they allow us to interoperate with developers and help humans understand the logic of how machines' cognition operates. And they change all the time - this is the treadmill we have no choice but to be on

Domain expertise (including AI). This means going both *deep* in a few specific spaces and *wide* in others to develop a broad-based generalism and an ability to see patterns across very different environments. This helps us critique AI suggestions and point machines at the right dots to connect. *Digital and other IT technology* expertise, of the deep one type, is one of the fastest-growing segments among domain expertise skills. Additionally, irrespective of what deep expertise we leverage for a living, we must stay current on what AI can do for specific tasks in those domains.

These are currently somewhat disjointed, if interrelated, disciplines. Perhaps we need a new federative one called "**augmented thinking**." The new curriculum should be designed to equip individuals *and* teams with the ability to sense, remember, create, decide, act, and learn effectively in complex environments. It should integrate foundational thinking skills, including logical reasoning, analytical skills, and reflective thinking, with cognitive flexibility, problem identification, solution generation, critical thinking, decision-making, implementation, and continuous improvement, fostering both individual and collective intelligence. The curriculum would emphasize creative thinking, ethical reasoning, information literacy, communication, collaboration, emotional intelligence, and adaptive learning.



This combination would natively address the opportunities and challenges of a tight synergy between humans and machines in the AI-enabled, individual and collective cognitive process.

Yet, we must recognize that very few things are certain, and it is important to prepare for uncertainty.

Two certainties

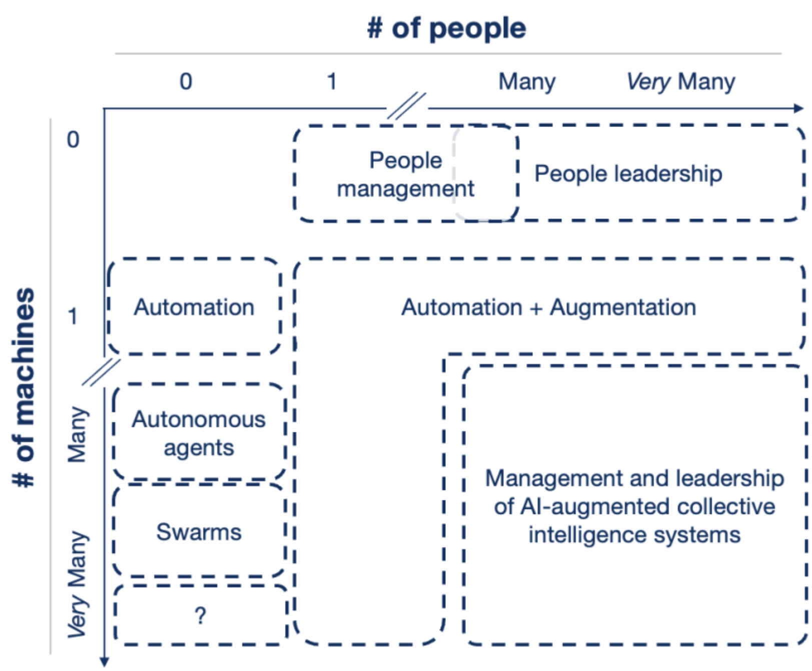
First, things will likely change radically in the long term, and the list above will evolve. But required capabilities and skills will likely hinge on

Human intelligence. Individual (self, e.g., resiliency, initiative, efficiency) and people (including organizations, network) leadership

Machine intelligence. Individual machine and machine-network leadership, i.e., designing automation workflows that use AI extensively, influencing the behavior of groups or even swarms of AI agents

Collective intelligence. A combination of all of the above, i.e., influencing the behavior of large groups of machines and people - aka superminds. For example, things like ACI’s four pillars: (1) Find or help find network (human or machine) **nodes**, that is, entities that participate in the collective cognition; (2) Give the right **incentives** to the nodes to collaborate, e.g., through culture shaping, sophisticated business cases, or change management. (3) Harvest the right **information** into the supermind to help it overcome its possible insularity and trigger the association of relevant new ideas (4) Help the supermind **collaborate** - e.g., through the right framework to solve specific problems such as strategy, debate, etc. or by using the right technology tools or methods.

The following chart is one possible representation of the future canvas of management and leadership, or, in other words, what is for us to work on. In each of the boxes, there are specific competencies for humans at the (a) design, (b) build, and (c) run stages - future jobs will fit into one (or multiple) of those.

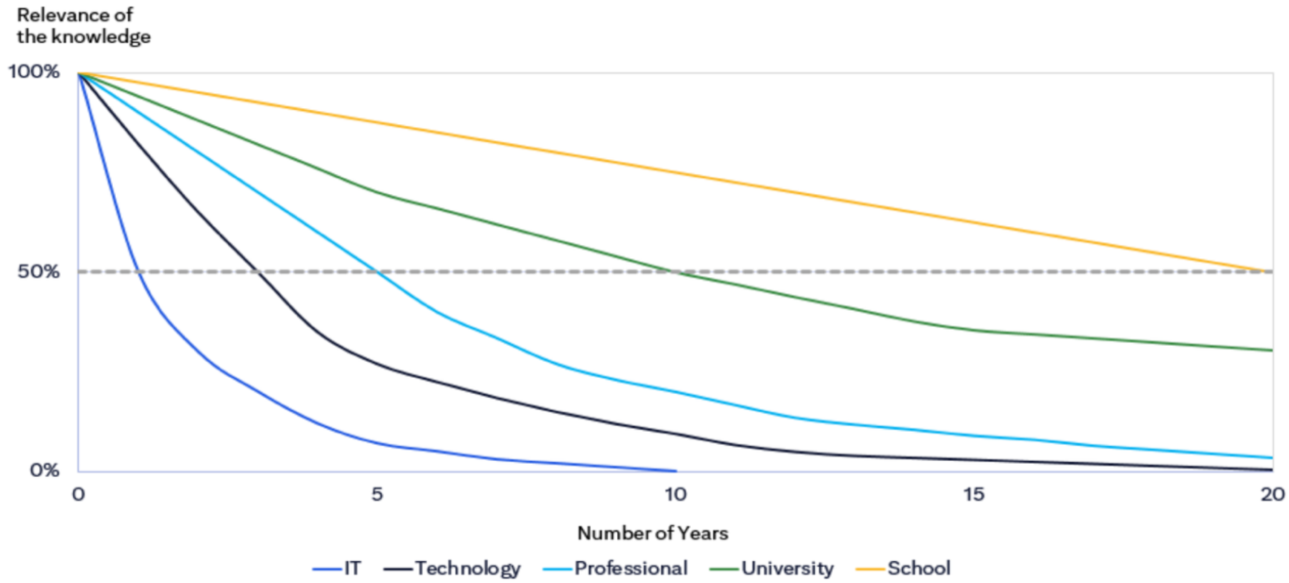


So, for instance, the skills required to *design* or *build* a workflow where multiple humans in the loops manage an agent in different process steps will be different from those required to *be* effective humans in the loop at run (or inference) time.

There is a second certainty. If this seems like a lot, it is because we need to learn from partially different disciplines and require a different, more agile, and faster approach to skill formation, which many companies today cannot provide. Given the pace of skill obsolescence (see chart below), this is not the right time to slow down innovation in skill infrastructure.

Let's conclude with an analogy. Builders *use* bulldozers—they don't outpower them. They need to orchestrate the interaction between physical space and bulldoze so it is tractable to the machine, including following specific design instructions and cooperating with tradespeople who work alongside them. For a while longer, a chunk of real-world tasks and quite a few

Figure 3. The half-life of knowledge



Source: Citi Global Insights

conceptual tasks will still need human support to make them tractable by machines. H[umans in the loop]() will continue to play an important role if we embrace it for what that is. Whether we want it or not, we humans are in the process of becoming mostly managers and even leaders with many assistants. Humans must direct the collective cognitive attention to the right things - the right "*whys*." We must ensure the approach is right - the right "*what*." We must critique the "*how*" that machines will increasingly suggest. These three steps require human-machine synergy, individually and in groups.

We don't yet know how to do that well. That's the learning charter for the foreseeable future—for individuals and organizations.

But let's be clear about one thing: we will spend a lot more time learning than we ever did before. Better get good at it—and get going now.

Three Skills Get GenAI to Do More For You

GenAI is currently overhyped as a *standalone* technology *and, at the same time,* much underappreciated as a complement to others and to us. To benefit from its full and evolving potential, I argued elsewhere that we need to learn what I call "augmented thinking," which encompasses several skills, including what we will discuss today: *the ability to guide* your AI (or multiple tools, including various types of AI, like ChatGPT, Claude, and Perplexity at the same time) and your team working on it toward the best results.

Effective use of GenAI is not just about typing queries and hoping they return accurate or insightful results. I believe the solution is not generic "prompt engineering"—at least not in its current scope.

Many believe that the future of prompt engineering will not be overly complex, especially for non-technical individuals. Models are increasingly adept at interpreting our requests, providing reassurance that adapting to AI technology will be manageable.

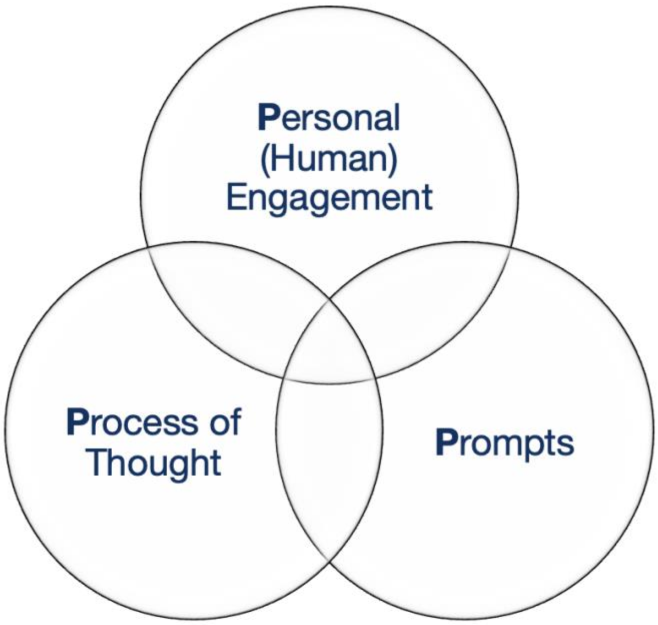
However, *learning to holistically engage more effectively* with GenAI tools does and will likely continue to yield superior results - and it's not a skill that most people have today. I see "Talking with machines" as part of tomorrow's core literacy, just like the human-engagement equivalent when we manage processes, group dynamics, and individual colleagues to get to better problem-solving and idea generation. We learn the human version of that over formative years of management practice, and much literature exists about it. Some people are good at it; others are not; everyone can learn some - especially to make the results more consistent. So what is the equivalent when managing the artificial workforce composed of indefatigable artificial "interns"? Everyone is a manager of these machines now - and we better become good at it.

Three lenses for harnessing GenAI tools: GenAI 3Ps

Three dimensions help us understand what needs to be learned. We will go over each.

1. **Personal, Human engagement**. Ensuring that humans are still firmly in the loop and not "asleep at the wheel."
2. **Process of Thought**. Leading the machine through a thinking process, a sequence of requests, data inputs, etc.

3. **Prompts, one by one**. Engaging optimally with the machine through better prompts (requests), one at a time.



They overlap, just like in the non-AI-enabled world, when doing the same with humans. Think of innovation workshops that use design thinking techniques: you need to manage the human factor (find the right people, motivate them, engage them), follow a process (have a directional sequence of steps, but also know when to change direction), and get the participants to perform individual activities (supported by artifacts for best results).

Let's learn how to harness them one at a time.

1. Engage with the technology

Let us quickly settle the "human in the loop" part, as I have written about it elsewhere and included references to work done by our MIT team and many other world-class researchers below. The key points are:

1. **It is a significant problem**. People do fall asleep at the wheel, becoming over-reliant on unreliable AI's skills, and when they do, quality suffers - both in terms of accuracy and creativity
2. **We often don't pinpoint the "where/when".** Part of the issue stems from the current inability of professionals to identify which subtasks fall within AI's frontier of the art of the possible and which subtasks are likely to be too repetitive for humans to maintain typical performance. The upshot is that people often give AI things that are too hard to do or don't use them to lighten the burden of repetition. More on this here.
3. **Designing the UI/UX** for people, including specific scaffolding of the human-computer interaction (more here), provides an exoskeleton that mitigates these issues
4. **Designing a scalable human-in-the-loop people/process/tech *stack*** is serious business; there is still much work to do in this space.

I argued in the past that we might need something similar to what happened when Six Sigma and Lean were introduced in work processes, which is to say, we need to expend significant time and effort (and brains) on it

5. **You drive**. A straightforward rule for our current individual use of generative AI is always to have the initiative—don't let the machine have it. That means aggressively explaining, critiquing, skeptically asking for proof of logic, asking to be asked, etc. A fully self-driving car has yet to be on the market, but assisted driving is already beneficial. Similarly, we should not treat current GenAI technology as a fully autonomous self-answering bot, especially for complex problem-solving.

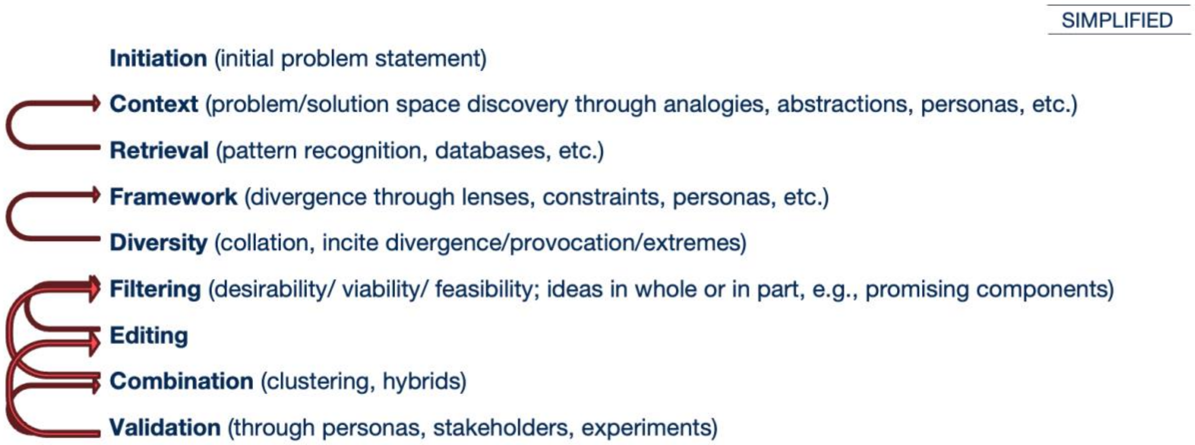
2. A process for your augmented thinking

GenAI can't think about everything all at once at inference time (as it responds to your query) and cannot know everything relevant to our circumstances. That's where a process is helpful.

Level set. First, today's GenAI tools mostly don't know who you are and what you're trying to do unless you explain it, just like what you would do with humans. We all - us and them - need context: data injected at the right time, remembered, and surfaced at the right places. Steps in a process help collect the relevant context and narrow the scope of the possible answers. Compared to humans, the difference in the case of GenAI is that if you use a chat (ChatGPT, Claude, Gemini), they lose context after a while, and you need to either remind them periodically and/or stay concise. Thankfully, context windows are becoming more extensive.

Be aware of computational efficiency limitations. Second, GenAI models are limited in the inference they can make in every run and in the number of tokens they use to be cost-effective. Like humans, you can help models by breaking the thought process into discrete steps.

Follow a (meta)process. There is much literature about ideation, but I suggest starting with this article and the chart below. The creativity meta-process is something that you and your teams have likely used already. You will want to follow that flow or derivatives, although recursions (back and forth, as opposed to a linear path) are always possible and often needed.



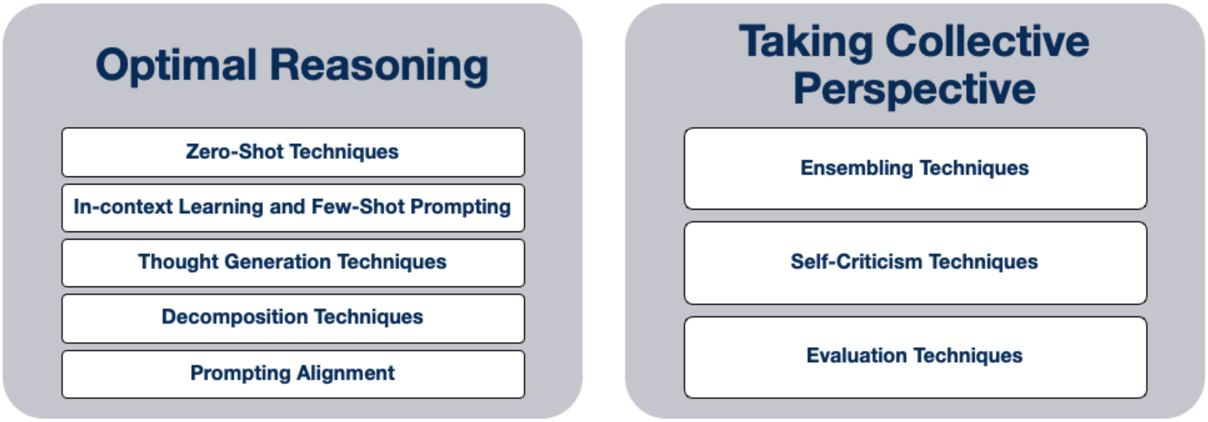
3. One prompt at a time now

Next, the prompts. (This topic occupies most of the public discussion, and for a good reason: prompt engineering is still an art when scaling things in software production, and developers want more predictability. However, here, we *focus on the application of these concepts to the work of non-technical* people. For them, prompting, in a narrow sense, is not the only important part of an augmented thinking skillset.)

There have recently been some excellent meta-analyses of all prompting techniques (see the paper “The Prompt Report: A Systematic Survey of Prompting Techniques” and Daniel Lopes' helpful blog "A Comprehensive Guide to Text Prompt Engineering Techniques"). I summarize some of their findings below and in the appendix, but I also categorize things slightly differently so we stay focused on augmented thinking instead of narrow prompting.

Many of these methods leverage collective intelligence concepts. They harness AI's ability to compare different viewpoints, whether from the same model at various steps (which I group under “optimal reasoning”) or from other models and benchmarks (“collective perspective”). That dialectic helps the emergence of new, strong ideas. AI researchers are latching on to related properties because they help machines become more effective. Our role, as humans, is to *engage with them to make that dialectic happen*.

There are many ways of doing that, and I group them into two clusters: optimal reasoning and taking a collective perspective.



Optimal Reasoning

- **Zero-Shot Techniques:** These use no prior examples to generate responses, but they may assign roles, styles, or emotions to guide the AI response (e.g., "You are a CFO reviewing this proposal"). Rewrite prompts for clarity. Encourage answers based on known facts of a character/persona
- **In-context Learning and Few-Shot Prompting:** They use a few examples to guide AI (e.g., showing "what good looks like"). The number and quality of examples may impact performance (the order may do so, too, though less so than in the past). Use optimal input formats when deriving examples from large datasets to generate optimal examples
- **Thought Generation Techniques:** They encourage AI to explain its reasoning step by step. Use thought-inducing phrases and structured formats like tables. (Note that for OpenAI o1, asking the model to think step-by-step is not recommended, as it will do it by itself.)
- **Decomposition Techniques:** They break down complex problems into smaller parts (e.g., "What are the parts or types of my problem?". Solve each part step-by-step or in parallel. Use both natural language and symbolic reasoning
- **Prompting Alignment:** These ensure that AI's output aligns with user intentions. Address AI's tendency to agree with users to avoid sycophancy, etc. Handle biases, cultural sensitivity, and ambiguous questions

Taking Collective Perspective

- **Ensembling Techniques:** They combine responses from multiple prompts (or models) for better accuracy and use specialized artificial experts (e.g., personas, simulating specific roles, skills, etc.) for different types of reasoning

- **Self-Criticism Techniques:** They help AI evaluate and improve its own responses. They provide feedback on its answers, make necessary corrections, and verify answers with related questions for consistency. They can also prompt the AI to ask the human(s) to evaluate its input further.
- **Evaluation Techniques:** They use structured prompts to evaluate text quality, employ multi-agent frameworks for diverse perspectives, and add automatic standardized steps (e.g., rating scales and definition) for a more thorough evaluation

The appendix below provides more details and examples, and I recommend that you identify some that you want to master and use repeatedly. Some are very complex for nontechnical use, but they can still inspire you. If you squint, you will see that many of these techniques and methods are constructs derived from classical logic (e.g., Socratic methods), metacognition (deliberate thinking about thinking), or other thinking methods (e.g., design thinking or lean) that you may already be familiar with. That should help your learning curve.

In the end, you will already derive much value from using your version of some of these techniques. What is critical is to focus on a few of these concepts deliberately—more in the appendix.

Conclusions: we, humans, manage GenAI labor

GenAI tools will improve, just like workers tend to do over their careers. However, just like many human workers, they need a proactive manager. That is you, me, and our teams. This requires learning skills that help us guide them as they augment our thought processes. In this essay, we focused on three aspects

1. keeping the human optimally engaged
2. ensuring a deliberate process of the conversation with the GenAI tools
3. talking to the tools in words that help them (and us) do the job more effectively

There's a learning curve in this. It is part of the new augmented thinking literacy that will harness the power of these technologies to give *us* superpowers. Experienced managers know that blaming their staff for poor results is pointless—learning how to guide them to the best results is much more effective. The same applies to our engagement with GenAI.

Appendix: an overview of the techniques that executives, not just coders, can leverage in their work

Group Name	Category Name	Technique Name	Technique Example and Description
Optimal Reasoning	Zero-Shot Techniques	Role Prompting	Assigns a specific role to the AI. Example: "You are a travel writer. Describe the experience of visiting Paris."
		Style Prompting	Specifies style, tone, or genre. Example: "Write a short sales pitch in a persuasive style."
		Emotion Prompting	Incorporates emotionally relevant phrases to enhance output quality. Example: "Explain the importance of this project; it is crucial for my career and personal growth."
		System 2 Attention (S2A)	Asks the AI to rewrite the prompt, focusing only on relevant information. Example: "Given this text, extract unbiased information and the main query."
		Simulation Theory of Mind	Focuses on the knowledge a character has and answers based on that. Example: "What does John know about the event? Answer only with what John knows."
		Rephrase and Respond (RaR)	Rephrases questions for clarity before generating the final answer.
		Re-reading (RE2)	Repeats the question to improve understanding.
		Self-Ask	Generates and answers follow-up questions before the final response.
	Few-Shot Prompting (Using Exemplars)	Key Concepts / in-Context Learning	Uses examples or instructions in the prompt. Example: "Translate English to French: English: Hello French: Bonjour"
		Few-Shot Prompting Design Decisions	Involves exemplar quantity, ordering, label distribution, label quality, format, and similarity to improve performance (some has become less relevant with new large models).
		K-Nearest Neighbor (KNN)	Selects similar examples for better performance. Example: Uses KNN to select exemplars close to the test sample.
		Vote-K	Ensures diversity in exemplars through a two-stage selection process.
		Self-Generated In-Context Learning	Generates exemplars automatically when training data is unavailable.
		Prompt Mining	Finds optimal prompt formats in large datasets.
	Thought Generation Techniques	Chain-of-Thought (CoT) Prompting	Encourages the AI to express its reasoning explicitly. Example: "Q: Jack has two baskets, each containing three balls. How many balls does Jack have in total? A: One basket contains 3 balls, so two baskets contain 3 * 2 = 6 balls."
		Zero-Shot-CoT	Uses thought-inducing phrases without exemplars. Example: "Let's think step by step."
		Step-Back Prompting	Asks high-level questions first before detailed reasoning.
		Analogical Prompting	Automatically generates analogical exemplars for reasoning.
		Thread-of-Thought (ThoT) Prompting	Uses an improved thought inducer: "Walk me through this context step by step, summarizing and analyzing as we go."
		Tabular Chain-of-Thought	Outputs reasoning in a structured table format, enhancing clarity and structure.
	Decomposition Techniques	Least-to-Most Prompting	Breaks problems into sub-problems and solves sequentially.
		Decomposed Prompting	Uses functions for different sub-problems and combines results.
		Plan-and-Solve Prompting	Plans and solves problems step-by-step. Example: "Let's first understand the problem and devise a plan to solve it. Then, let's carry out the plan step by step."
		Tree-of-Thought (ToT)	Uses a tree-like structure for generating multiple possible steps.
		Recursion-of-Thought	Solves sub-problems recursively and inserts answers back into the original problem.
		Faithful Chain-of-Thought	Combines natural language reasoning with symbolic reasoning (e.g., Python). Example: Generates reasoning including both narrative explanation and code snippets.
		Skeleton-of-Thought	Creates sub-problems and solves them in parallel for faster answers.
	Alignment	Prompt Sensitivity	Input prompts significantly affect output. Example: Different phrasings of the same question can lead to varying answers.
		Overconfidence and Calibration	Techniques for accurate confidence levels. Example: "How confident are you from 1 to 10?"
		Sycophancy	AI's tendency to agree with user inputs. E.g.: When the user implies a preference, the AI might align its answer to that.
		Biases and Stereotypes	Methods to ensure fairness and cultural sensitivity.
		Ambiguity	Techniques for handling ambiguous questions.
Taking Collective Perspective	Ensembling Techniques	Demonstration Ensembling	Uses multiple prompts with different exemplars and aggregates outputs. Example: Combines responses from various prompts to ensure a comprehensive final answer.
		Mixture of Reasoning Experts	Specializes prompts for different reasoning types and selects the best answer. Example: Uses different experts for logical reasoning, factual accuracy, and creative thinking, then selects the most accurate output.
		Max Mutual Information Method	Selects optimal templates that maximize mutual information between the prompt and AI's outputs, enhancing relevance and accuracy.
		Self-Consistency	Samples multiple reasoning paths for consistency (also using templates).
		Meta-Reasoning over Multiple CoTs	Combines multiple chains of thought for a final answer.
		DiVeRSe	Uses diverse prompts and self-consistency for scoring.
		Consistency-based Self-adaptive Prompting	Selects exemplars based on agreement and performs self-consistency.
		Universal Self-Adaptive Prompting	Uses unlabeled data for exemplars and a scoring function to select exemplars.
		Prompt Paraphrasing	Augments data by rephrasing prompts.
	Self-Criticism Techniques	Self-Calibration	Evaluates confidence levels of answers. Example: "How confident are you from 1 to 10?"
		Self-Refine	Provides feedback and improves answers iteratively.
		Reversing Chain-of-Thought (RCoT)	Compares original and reconstructed problems to check for inconsistencies.
		Self-Verification	Scores solutions based on prediction confidence.
		Chain-of-Verification (COVE)	Verifies answers with related questions and revises the answer.
		Cumulative Reasoning	Evaluates and accepts or rejects reasoning steps iteratively.
	Evaluation Techniques	LLM-EVAL	Uses a schema for evaluation with a single prompt. Example: Scores a text on grammar and relevance.
		G-EVAL	Adds AutoCoT steps to evaluation.
		ChatEval	Uses multi-agent debate for diverse evaluations.

Source: Analysis derived from the paper "The Prompt Report: A Systematic Survey of Prompting Techniques" and Daniel Lopes's "A Comprehensive Guide to Text Prompt Engineering Techniques."

Superhuman Knowledge Workers? AI Exoskeletons and Scaffoldings

To control, harness, and compete with AI (and AI-powered organizations), we need to make *people* better - individually and in their organizations, through their work processes. Even with today's technology and practices, there are signs that AI can help do that with knowledge work, as it serves as an augmentation technology when used well.

Some call the result of this evolution cyborgs, reflecting a tightly integrated approach to human-AI collaboration, with continuous, synchronous exchanges at the subtask level. That is opposed to centaurs, i.e., a model of human-AI collaboration where there is a clear division of labor between human and AI capabilities, and often, one set of tasks gets fully automated, and collaboration often happens somewhat asynchronously. Both are useful and will happen. This article focuses on the former: what they are, why they matter, and how to build them.

Early promise

Recent empirical studies underscore how Generative AI-based tools have the potential to enhance productivity and work quality across professional domains, but not by default.

Early studies show that AI tools like GitHub Copilot and GPT-4 can significantly boost productivity across fields, enabling faster task completion, higher quality outputs, and increased creative performance. Developers, customer support agents, business professionals, and consultants reported notable efficiency and quality improvements, handling more tasks in less time while enhancing overall output standards. Generative AI can also make individual contributors, especially those with lower skills, less dependent on others (and more efficient) and able to explore with more tools.

Yet, augmentation still has risks, and collaboration between humans and machines doesn't automatically result in better outcomes than each would achieve individually, as an extensive recent analysis showed.

How do we design and build tools to get human augmentation right? To answer that, let's first clarify what kinds of supporting tools exist.

What are AI exoskeletons and scaffolding for people?

First, let's clarify what scaffolding and exoskeletons in AI mean.

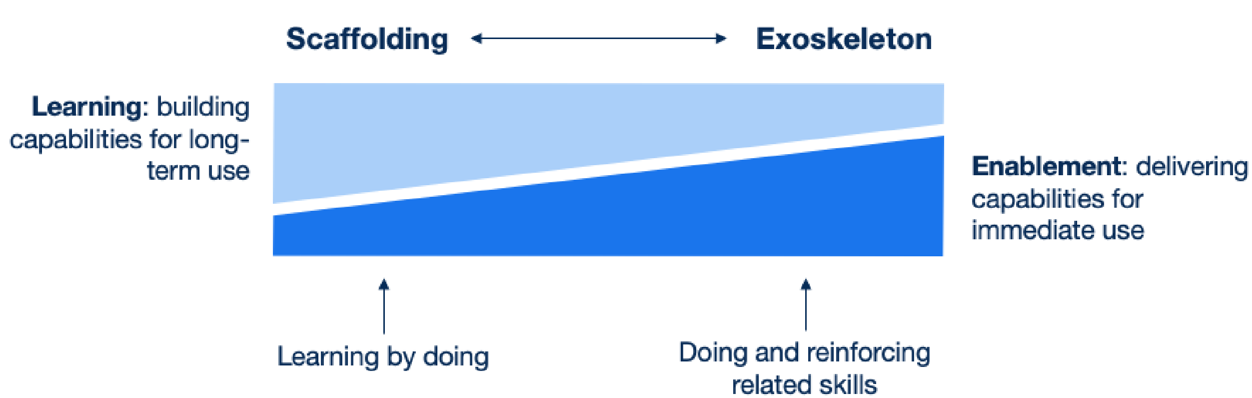
Scaffolding is primarily designed as a temporary support, focusing on cognitive assistance and learning processes. It is designed to encourage a more reflective and iterative approach to problem-solving. It adapts dynamically to user progress, gradually reducing support as proficiency increases, and aims to foster independence. Scaffolding is commonly applied in educational technology, software development assistance, and guided data analysis, targeting specific tasks or learning objectives. For example, an AI that shortens the initial learning curve on any topic or asks questions to develop specific critical skills.

In contrast, AI **exoskeletons** are more persistent enhancement systems. They aim for continuous augmentation of both cognitive and analytical capabilities across various domains without necessarily intending to be removed over time. Exoskeletons seek a symbiotic integration with human cognition and decision-making processes, providing ongoing support in real-time information processing, language tasks, complex modeling, etc. While scaffolding works towards user autonomy, exoskeletons maintain constant augmentation, becoming integral to the user's enhanced capabilities—for example, an AI "whisperer" or a problem-solving GPT.

They represent distinct approaches to augmenting human capabilities but sit on a continuum. Some scaffolding can also be used as ongoing support for execution, and some exoskeletons can help people learn.

How to build them right

The first clear realization is that we must “lean into the 70”, that is, 70% of the AI transformation work that isn’t about the technical, core-IT side - process and people. AI can help make them better. And most organizations don’t allocate enough resources to do that. While much of the headlines these days are about *autonomous* AI agents, the lower-hanging fruit is the design of the interaction of people, individually and collectively, in workflows.



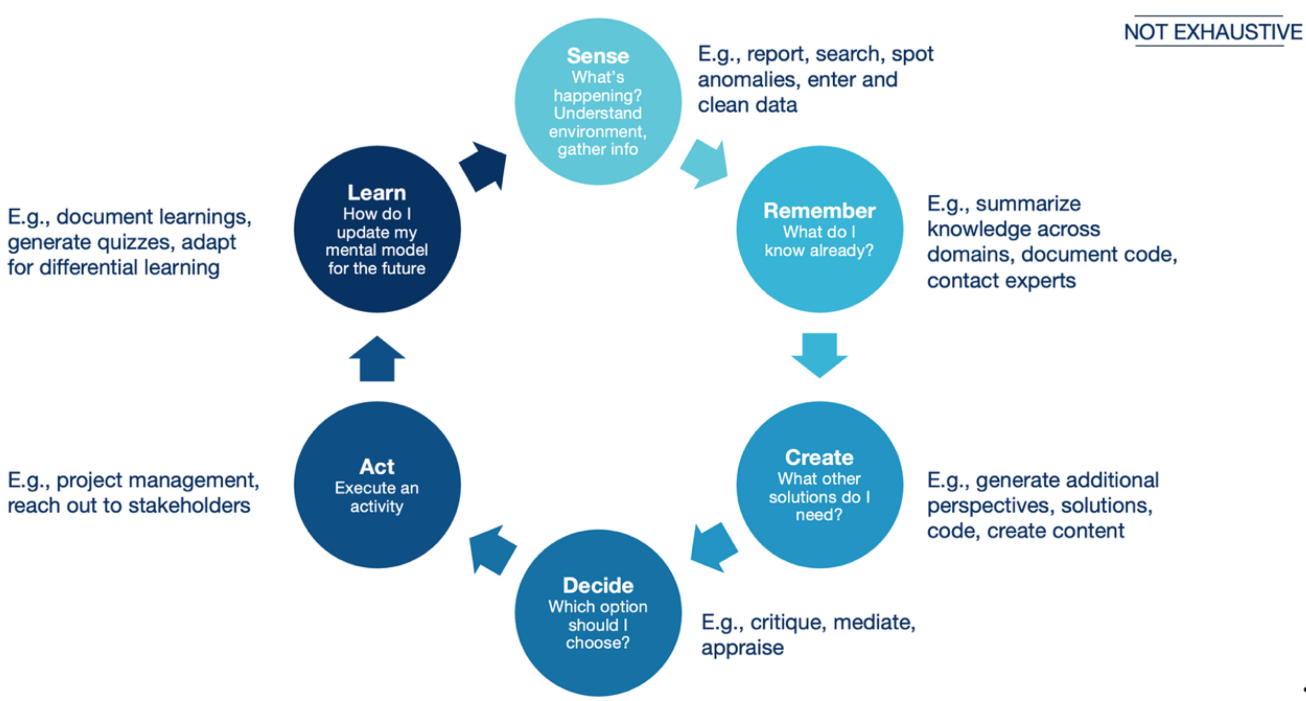
This means focusing on people's *capabilities*, including what I call Augmented Intelligence, which is the ability to work with machines seamlessly and act as a "System 2" (slow, abstract) to AI's "System 1" (fast, recognizing data patterns). In this article, however, I want to talk about the design of the *workflows* that help people achieve more, and better, especially but not only the problem-solving and innovation process.

To start with, building scaffoldings and exoskeletons right requires understanding how to **guide AI** along an appropriate cognitive process (more on this here and here).

It also means focusing on which AI capabilities are already good enough to improve the human ones. The framework below shows the types of activities people and groups perform (called "supermind cognitive processes"), hints at which GenAI is good at, and helps understand the potential scope of intervention.

Some experimental evidence from prototypes

Our research at MIT's Center for Collective Intelligence started in 2020 and has since shown the value of designing scaffoldings and exoskeletons well. Our latest study, for instance, compared the effectiveness of MIT Supermind Ideator, an AI-driven idea-generation and problem-solving tool we built, to ChatGPT and solo human effort. In addition to an intentional process reflected in specific prompts, the Ideator partially operates on a fine-tuned large



language model (OpenAI's latest) trained on case studies of innovative organizational practices. This makes it more context-aware and responsive to domain-specific challenges than general-purpose tools.

We found that individuals using it generated significantly more innovative ideas and engaged in deeper interactions than those using ChatGPT or working independently. The findings underline the advantage of specialized AI interfaces for improving human-AI collaboration in targeted, complex problem-solving contexts. By offering structured steps and prompts, Ideator is a scaffolding tool where people learn how to perform specific tasks optimally—and might even be an exoskeleton, usable constantly—that helps users overcome cognitive biases and explore alternative perspectives.

Interestingly, the data shows that AI tools like Ideator might be especially beneficial for less naturally creative people, at least initially. The study also emphasizes that, despite AI's support, human judgment remains essential for curating and refining AI-generated ideas to reach optimal solutions. It also suggests the need for designing the appropriate turn-taking and sequencing of human-machine interventions to avoid drifting into platitudes and risking dumbing down people over time.

Turning it into real-world impact

Combining domain expertise (business, innovation practices) with easy-to-build and easy-to-iterate AI tools has great potential. Low- and no-code AI technologies (like OpenAI custom GPTs, Langchain, or Wordware, to mention some) will eventually do what Excel did to all of us thirty years ago: allow people to translate business logic into repeatable actions, hence forever changing how work is done.

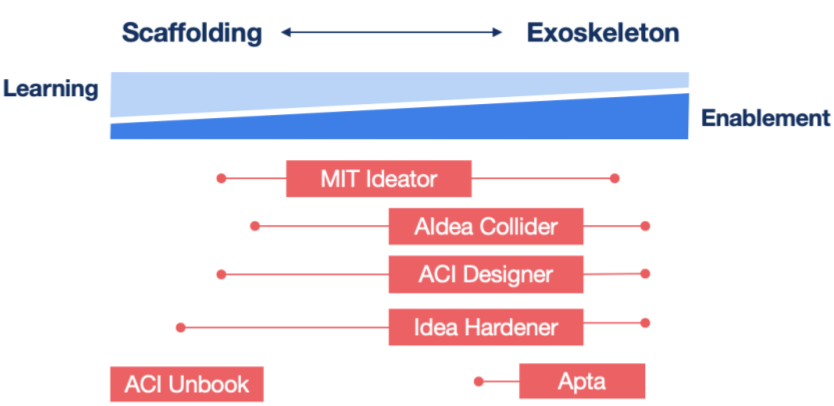
This is especially important as the "art of the possible" (the "jagged frontier") keeps moving, because continuous experimentation from laypeople, not just technologists, will drive identification of use cases and iteration toward viable product-market fit.

As an example, inspired by the above ideas, it was easy to encapsulate business processes and practices and curate knowledge bases (e.g., an extensive database of organizational design and process examples) into a few tools that help with problem exploration, solution discovery, and solution refinement. Some examples are below, all built on the theOpenAI GPT Store and freely accessible (find them through the store's search menu).

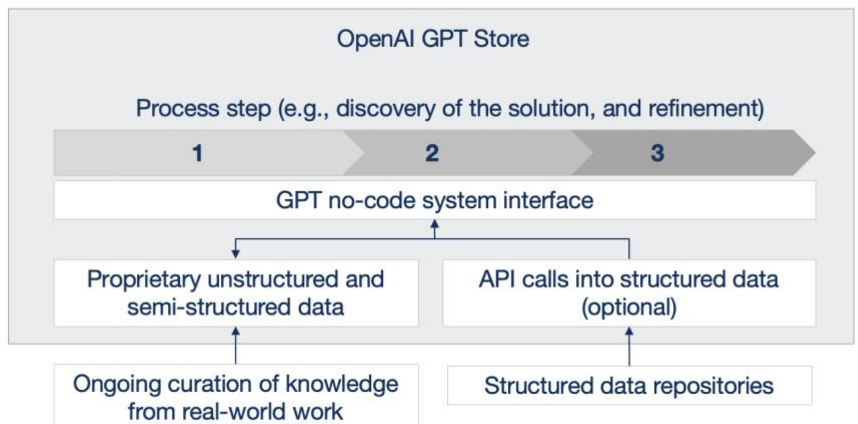
- **AIdea Collider** - thoroughly explores business and organizational problems, then systematically combines diverse solutions for superior ideation.

- **ACI Designer** - generates uncommon ideas for complex problems by leveraging ACI (Augmented Collective Intelligence) principles to support the design of collective intelligence (supermind) architectures.
- **Idea Hardener** - strengthens and refines ideas and concepts by using AI to harness dozens of critique frameworks.
- **Apta** - identifies parts of an organizational process that Predictive and Generative AI can augment or automate.
- **ACI UnBook** - guides the reader to discover the most relevant concepts from a set of long-form content (a book) and other fact-bases

To an extent, these can be used as scaffoldings and exoskeletons, depending on users' preferences. Less-experienced people who rarely do those tasks can use the tools to perform at a "better-than-beginner" level. Those who want to permanently flex that muscle more can use the tools to learn faster. And skilled people use the tools to complement their work.



Illustratively highlighted below, their architecture is straightforward. Non-technical subject matter experts can now design and even build some of them.



This hints at the possibilities, especially when focusing on specific personas (e.g., the CFO in a board meeting or a salesperson preparing a proposal) and adding specialized knowledge that organizations routinely accumulate. Think of use cases like:

- **Strategy alignment.** Translating company strategy documents into "Chief Strategy Officer's digital twins" tools enables strategy socialization, cascading, and team alignment. It also allows leaders and, in general, all employees to constantly query the strategy without needing access to the strategy team.

- **Update of staff knowledge.** Learning anything, not just new skills, in a more agile way - for instance, helping technology implementation teams understand how new products are better than old ones or bringing people up to speed on and contributing to new use cases for GenAI.
- **Avoiding blind spots**. Knowledge "whisperers" will be curation agents that incessantly scan within and outside the organization to find relevant and non-duplicative knowledge (compared to their human users' knowledge base) and summarize it for the user daily or whenever they perceive the human is trying to decide something related.
- **Board prep.** Assisting CXOs in preparing for board meetings by buttressing their thought process and preparing for objections.

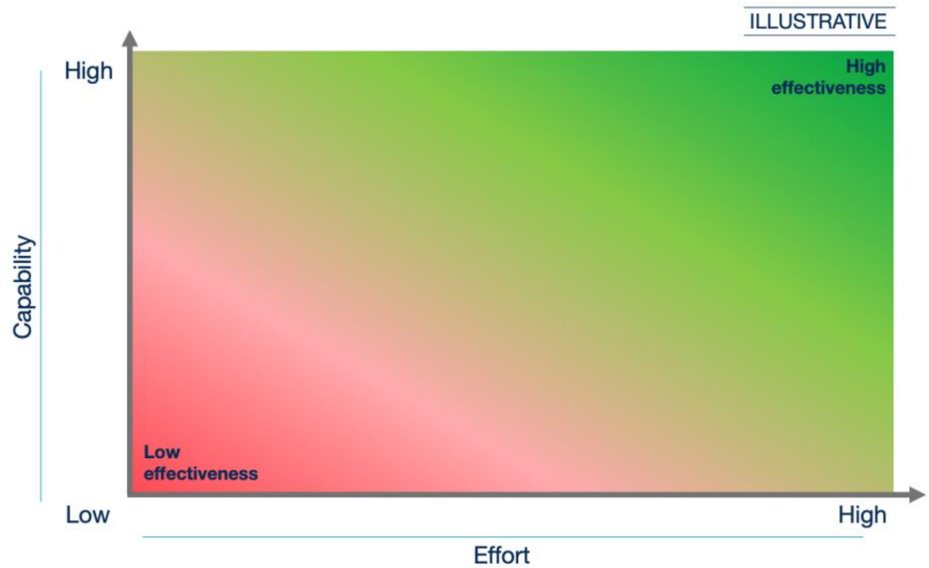
Many others exist, both horizontal across roles and even more excitingly vertical, role-specific and domain-driven ones.

Like Excel forty years ago, which is now used by anyone (and much of today's organizations, willing or not, run on Excel and its macros), these AI-powered "mini-apps" could generate significant productivity improvements and alleviate employees' toil. And they might do something else: through widespread, bottom-up experimentation, they might surface use cases that don't exist today. And, of course, we will increasingly use more sophisticated tooling, all the way to AI agents or even networks (graphs) of them - with a human in the (very tight) loop.

The upshot can be more-effective people, enjoying their work more. But today, in most organizations, there is an imagination, not just a technical gap. That gap can be closed. It is time to start designing and building AI-powered exoskeletons and scaffoldings.

Capability + Effort: Exploring AI's Jagged Frontier

Much of the discussion on how to use AI - specifically Generative AI - is about finding appropriate use cases. Progress has been made in the last months on that front. In that context, a useful idea is that of the "jagged frontier", highlighting how artificial intelligence can be more or less effective in individual tasks than qualified humans. Exploring that concept can benefit from a deeper look at how people do things.



It is not just about capabilities

The frontier is not static. Not just because technology advances, or because of the improvement of processes that harness it, but also because it depends on each of us - our capabilities, and our *moment-to-moment* states - or more specifically, the effort we expend on each task, at a specific time.

Let's take an example. Imagine some illustrative (but not made-up) steps a knowledge worker takes to generate and syndicate a recommended solution for a problem. We can analyze those steps based on two dimensions: the person's capabilities, and the amount of effort s/he (or it) puts in. The effectiveness of the outcome depends on both.

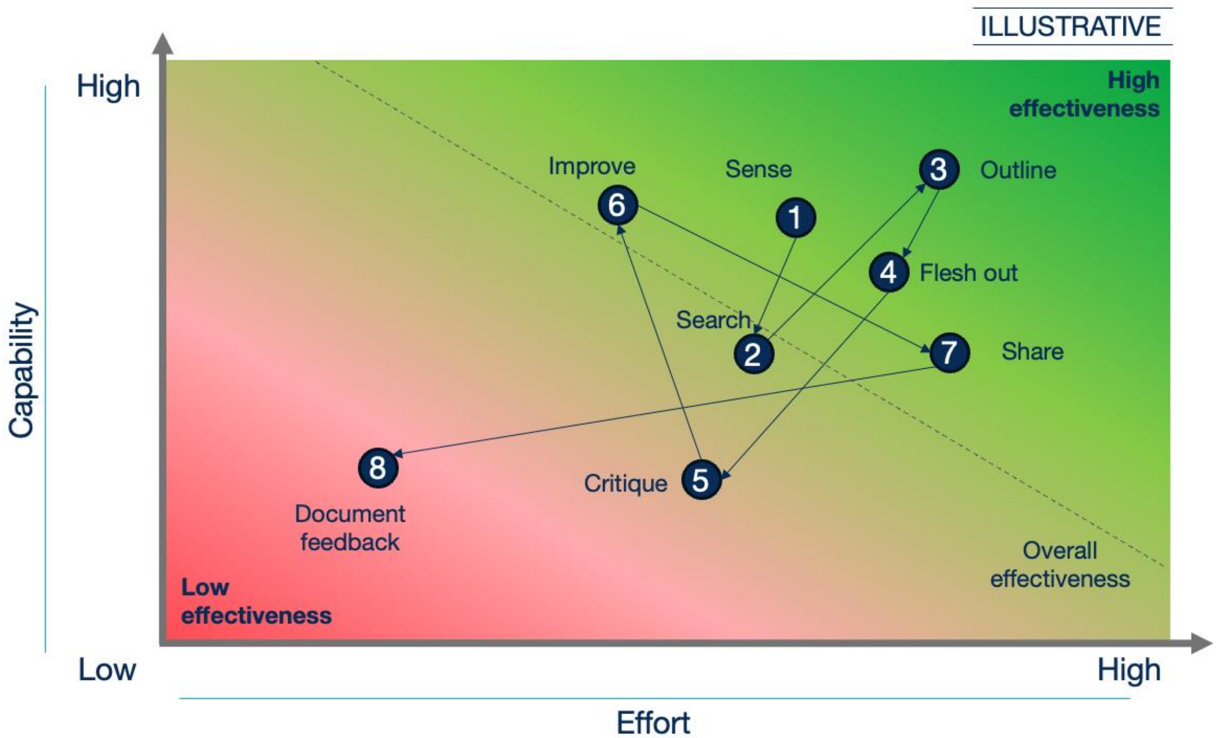
The steps, in the problem-solving example, would broadly fall into the following buckets (adapted from Malone).

(1) Sense e.g., interpret the needs of the organization, and identify the gap compared to what's possible

(2) Remember if there's already a solution, search for it in the knowledge management systems, or reach out to colleagues who might have parts of it

Create a new solution: (3) outline ideas, (4) flesh them out (5) critique them alone and then in a group (e.g. with colleagues), possibly integrating perspectives from other fields, (6) then improve the documentation, and (7) share it

(8) Learn from the feedback, both individually and collectively, e.g. by documenting the responses



Each of us is good at some steps, and we like some more than others (say, outlining, instead of receiving hard feedback and documenting things thoroughly). We also get tired after a while and tend to cut short things we struggle with. Our quality of output is, as a result, uneven.

Those are all toeholds for AI to help through its capabilities, and its *capacity* - that is, its ability to expend effort on anything.

How does AI improve these?

Some examples:

Sense - monitor enterprise-wide conversations, e.g., through summaries of threads (Microsoft Teams, Slack, Notion) or transcripts of calls; feed ideas from external sources, revealing a possible gap compared with the art of the possible and hence an opportunity

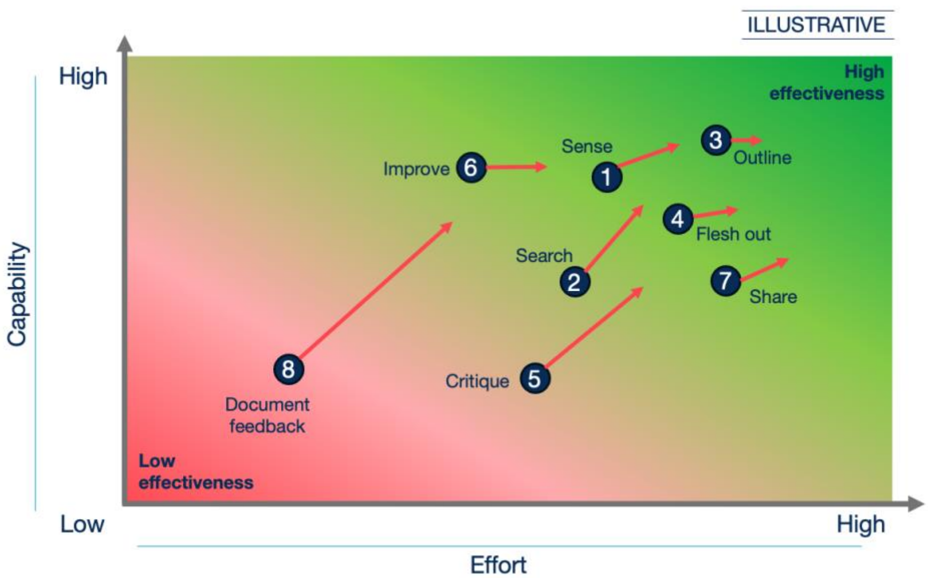
Remember - search the existing organizational knowledge, facilitate the identification of subject matter experts who could be contacted

Create—AI is less adept than expert humans at applying symbolic, conceptual thinking outside the box and might generate platitudes if left to its own devices. However, it is often very good at breaking down the problem, finding analogies, critiquing solutions from various perspectives, creating stories for different audiences, etc.

Learn—AI can memorialize, summarize, and store the documentation derived from the downstream conversations about and even the application of the idea.

This is thanks to AI's capabilities *and* effort - as AI is virtually indefatigable if prompted to be so.

The red arrows in the chart below illustrate this effect.



Arrows pointing up indicate that AI can already augment the capability of most people (e.g., by accessing new knowledge sources), and arrows pointing sideways mean that it enhances the *capacity* of the human counterpart (for instance by doing the same job multiple times in slightly different ways).

The outcome can look like the last chart below. The blue numbers are the final, augmented steps, the greyed ones are the initial ones, and the dotted lines show the overall improvement ("iso-effectiveness") levels.

This is just an illustration, but it helps deconstruct the human experience as part of a process and understand where to focus our new tools and methods.

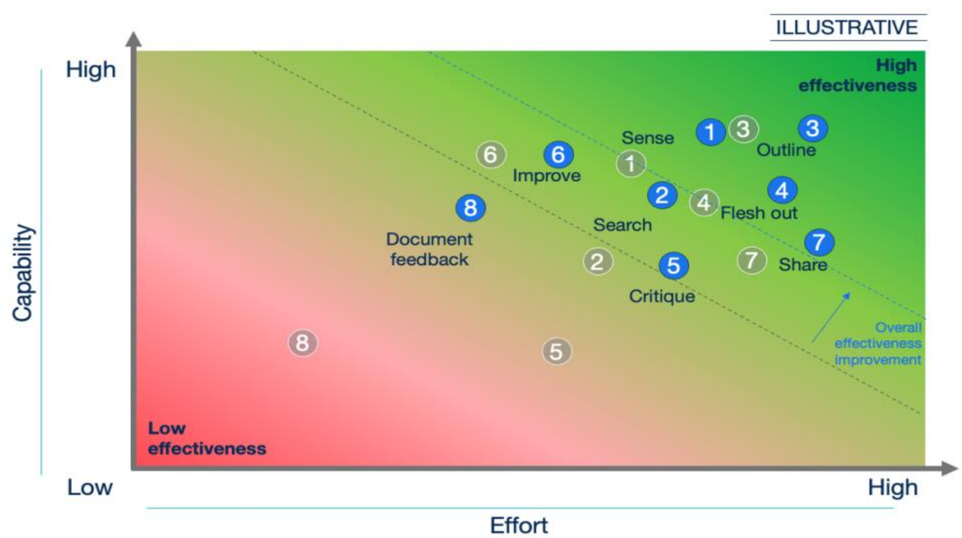
Time to design for augmentation

There are at least three implications.

First, **individual professionals** must experiment with AI and find what works for them, given their capabilities and attitudes. That is, they must attempt to use AI to mitigate their "low effort" tendency with some types of tasks.

The framework is also useful for understanding the **possibilities of** process improvement, alongside more traditional methods (e.g., experience design, and Lean, among others).

Organizationally, the upshot is that leaders must facilitate employees' update



of these new capabilities, both with current individual productivity tools and with more sophisticated process designs. And in the future, with more autonomous, agent-driven processing.

This analysis merely scratches the surface and can be expanded. There's a rich canvas for designing a future of work where AI supports our individual and collective intelligence by enhancing *both* our capability and capacity.

On AI's Dislocation of Human Labor, We Owe The Workforce Better Guidance

This is an informal review of research (then-current as of the end of 2023, and widely relevant as of the latest revision of this paper) on GenAI and talent strategy, and I draw a few conclusions I haven't seen being made explicitly enough.

This is not an HR-only conversation. With AI, in a strange way, we are all in HR right now.

Three topics I will cover:

1. Future work disruption requires guidance - HR and business counterparts need to collaborate to support the workforce
2. There are surprising results on "future skills" for humans to learn - there is no sacred ground, and the ground is shifting
3. Skills-building guidance and delivery - we need a better planning mechanism and doing corporate "tick in the box" is dangerous at many levels

1. Future work dislocation has been announced. Workers are drawing their own conclusions.

First - AI is now finally poised to change the world pervasively - if we are smart, for the better (and we do need that). Gen AI specifically can change our world in ways we always wanted AI to. Many workers are also excited about removing drudgery. But many are also **concerned**. And they're already showing signs of restlessness.

People have been told that there will be dislocation soon, but they have not been told how to address it.

If the implication of this feels hard to grasp, think about a similar situation that you likely know better: imagine there's a vast reorganization of your company, some people will lose their jobs, others will be moved around – but you are yet to receive the details, and it might take months, possibly years before you know. Things will likely trickle or come randomly out of the blue. How would you and your colleagues feel? That's the feeling.

To be sure, **if workers don't participate in the benefits, they will obstruct progress because they will not perceive it as such.** For instance:

1. they will not surface promising **use cases**
2. they will sandbag **change initiatives**
3. they will force HR to make all sorts of possibly unsustainable **concessions**
4. and there will be **volatility** – unrest, even.

That's a gap we need to fill - especially because **we are not ready to plug-and-play AI yet**, and there could be a period when talent is restless, and we can't deliver the automation/augmentation we need to do the work. The time to think about this is now.

Several data points are below. Starting with some data from PWC, pointing out the perceived positives expressed by employees. Similar data came from studies of software developers early in the year, as well as the Society for Human Resource Management who pointed out that the majority of people hope generative AI will eliminate some of the drudgery in their work.

Unfortunately, this perspective is dwarfed by the data from other studies. Let's also keep in mind that humans are primed to pay more attention to threats, and amplify them - an effect the media then latches Much data shows an interesting regional breakdown, and some data points indicating a different view of the regional concerns (developing the economies workers are generally more positive about it).

2. Surprising results on "future skills" for humans. No sacred ground - and the ground is also shifting.

So what is the defensible scope of work for humans? We used to believe that "humans will use empathy and creativity, and machines will do the scaling". We thought those skills were a human bastion. That is not true anymore. **There is nothing intuitive about what people will do**.

Machines can do a lot of creative work, and not just visual arts – this is the core of the work my team at MIT does, among others. But also, and even more surprisingly, machines can do empathy work, especially with their multimodal capabilities. More in detail:

Synthetic empathy is real. Humans seem to prefer interacting with a machine for healthcare interactions, under the right conditions (see "Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum"). Machines can already detect visual interactions between people and determine mood, and intentions, including cultural faux pas

Creativity and innovation have a new player, who wants to play with you. New studies show that part of that process is very effectively supported by machines, that...

a. do a really good job at exploring the problem space (see our MIT work on Ideator – and in general on AI-augmented, distributed second brains, aka superminds, and the Harvard/BCG study mentioned elsewhere)

b. produce more and better ideas than most humans, with lower variance of quality (for instance, see Wharton's

"Ideas Are Dimes A Dozen", from which the below chart is excerpted)

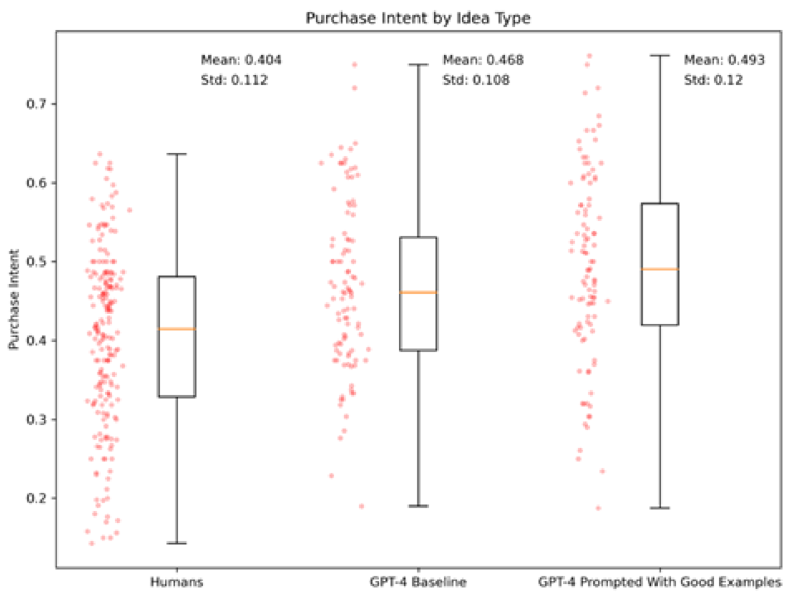


Figure 1 - Distribution of idea quality for three sets of ideas. Purchase intent is the weighted average of the five-box response scale per Jameson and Bass (1989).

3. We must give people *specific* guidance and support for jobs and skills

But that's not easy, because apart from the macro uncertainty about growth, we don't know what machines do well (the frontier changes), we don't know what the humans do well with the machines, and most of our workforce planning and training methods are somewhat obsolete, slow, and not granular enough.

They also don't keep into account that replacement by machines, which means that **humans will not learn some basic competencies anymor**e, because they will not do certain jobs anymore (some entry-level work - such as basic data analysis, etc).

I will be talking about (1) jobs needed (2) parts of those jobs that will still be human for a bit (3) how to get people there. Right now, there is too much noise from media headlines, and a lot of it feels like handwaving.

Workforce planning for AI is a real thing, and we must start with the jobs needed, not with the skills.

While performing it though, remember that the **possible frontier of the automatable (machine-human, mediated by processes) moves!**

Entire jobs will typically not be wiped out, but the structure of their tasks will undergo significant changes. Some tasks will be taken by machine – those within the frontier of the art of the possible.

We do need new frameworks and processes that help us to continuously assess

- what is within the "AI frontier",
- how jobs' *task structures* change over time. For instance,

salespeople will be spending less time writing emails, or coders doing a lot less documentation,

- and what to teach people so they move out of within the "AI frontier" and into *new* tasks that they can do and machines can't - for instance, pulling together genuinely novel analyses that don't feel like empty management calories.

In particular, we need to start thinking about "**skill stock",** not just people and jobs. Like the stock of semi-finished materials when you manufacture products. Based on

- **growth of demand** for specific jobs (e.g. we will code more, companies will become more fact-driven), human supply availability e.g. we will likely not produce/hire many more data scientists with [insert your industry domain] skills and we will lose people (because of retirement and/or attrition)
- **minus shrinkage** of the need for humans: **automation** of some tasks, and the fact that we will **democratize** the use of tools e.g. through Code Interpreter and other Copilot-type tools, which means that some of the existing load will be taken over by others, less specialized people

This algorithm can help us approximate the skills and the amounts of them we need to focus on.

With this in mind, consider that people will mostly be useful if they can work *with* machines, as they turn into, in the words of BCG and Harvard, "centaurs" and "cyborgs". But can they do that?

For now, research is mixed (see for instance "Generative AI at Work" from earlier this year) and shows differential reactions (some good, some bad) on employees' interaction with machines

- "Autopilot" for some, especially the less-skilled (they let the machines make all the decisions) as shown by the mentioned research from BCG in consulting, and on defect-detection on circuit boards manufacturing ("Lean back or lean in? Exploring social loafing in human-robot teams")
- "Experts' refusal" (they do not trust the machines) There is also a third reaction – the thoughtful and productive use of AI-augmented intelligence. That is where we want to lead organizations. We should call out this third option.

...which means that to stay relevant, people will need to acquire certain traits. What can we do about it? Among others:

1. **Generate awareness** at different levels of employee expertise and performance; generate awareness

about the performance level of AI in specific workflows.

2. **Build workflows** that embed prompts for people to pay attention to and probe the machine. This could be UX design for instance – such as in customer support where it is comparatively straightforward.

There's a gap that we should "teach into".

Then, skills building is a big part of the answer.

A few recommendations:

a. Train properly, no lip service: the BCG/Harvard study shows that “the appearance of training” may give people overconfidence and increase error rates when dealing with AI.

b. Train “irrespective”. Whatever the skill, people need to deliberately learn some of the aspects of the business that they used to “learn by doing”. For instance, the meaning and limits of statistical analyses, and even some of the basics of data wrangling (even if AI will increasingly do them). It is like being promoted to manager without ever having worked at ground level, which may accelerate the occurrence of Peter Principle's instances (that is, becoming incompetent in the new job). **We might need to train “in-silico” what we trained in real life**.

c. It is NOT just about prompt engineering, at least not beyond the basics, and not in the long term. Machines will become good at understanding what we want, in detail. The skills to interoperate with the machines will be more sophisticated, both at the Design/build phase (say, building models and UI/UX that interfaces them with the world) and at the Run phase (see points below)

d. Foundational skills matter. Some ideas:

Collaboration, project management, change management, critical thinking (the latter is in my opinion the uber-prompt engineering…as we need the questions very clear in our own heads first, or else the machine will be a rudderless speedboat)**;**

Related need for “interpretation” of models’ output; “explainability” of models will likely be part machine, part humans probing the machines e.g. triangulation, logic.

e. Everyone’s manager now that they have machines working for them. Being a manager is a special skill set, especially for a manager of this type of digital labor, who will need critical thinking, for instance. Revisit people leadership and management skills – now with machines in the loop, and identify the gaps. For instance, triangulation of sources and input is huge, as well as framing the questions, and sequencing and verification of outputs.

f. This leads us to the last point: Managers need to do a **sort of new-age Lean Six Sigma** (quality control, process improvement) on their processes again. This is digital labor prone to defects, just like people have always been (!). Start with the classics. More on this here.

Don't wait for your Human Resources department to do this - everyone is HR here

The science and practice are being written on this. We can write it collectively, across technology, operations, strategy, and HR. Failing to do this means keeping people in a very uncomfortable darkness - and what goes around often comes around. We need AI to help us, but much of the real work is on our human workforce. 🧠

With AI, Learning And Reskilling ≠ Training.

We need people to harness AI's power - but today, "out of the box," neither AI nor people are ready to interoperate seamlessly. Apart from generative AI's accuracy or ethical challenges, the broader question is about *designing and building the new work* - the new processes that optimally apply AI's capability *in the flow* of human work. Today, we don't have enough people to achieve that.

This results in:

1. widespread **inappropriate use of AI**, for the wrong things, i.e. deploying it on the wrong side of the "jagged frontier" some have talked about - for instance, trying to make it work in isolation for topics that require accurate, fact-based answers
2. **disproportionate focus on automation and displacement** of humans (the "Turing trap") instead of a more holistic focus on extending / augmenting human capabilities both as individuals and collectively in teams, organizations, and ecosystems.

Intuition, and pattern recognition based on our experience, also don't fully help.

- **Some of the current AI challenges are similar to what we had** with previous (predictive, machine-learning-focused) AI waves: faster company and *personal* obsolescence. The need for cross-disciplinary capabilities, including AI and the rest of data/digital, as well as domain expertise, process, and human-centered design, is still unmet.
- **However, part of the new AI wave is different**: generative AI is akin to the introduction of Excel at work, as it is implementable by everyone everywhere through their "copilots". Much of the usage will happen decentrally, through millions of little experiments where humans use the technology as a productivity-enhancing assistant. Today's Generative AI can make us much smarter, extend our capabilities (for instance, access to knowledge), or dull us (e.g., making us dependent on the machines' logic). This has a broader implication on what humans will do, and how much of the economic profits they can capture - to illustrate with an example, there are only so many people who want to annotate data sets, and that job doesn't pay that well. The other reason this AI wave is novel is that it is both fast and somewhat unpredictable. We

> haven't yet fully understood which tasks will be taken over by AI-enabled tools, but we know that *today's answers will be subject to change*.

While most try to figure things out, workforces become restless, and the best competitors take the lead.

As a result, one thing is certain: **we need an up/reskilling architecture that can adapt fast**, both in terms of what skills are taught, and how teaching is performed. Most organizations don't have that yet.

Up/Reskilling practices aren't fit for this purpose yet

Many existing skill-building architectures are not fully fit for this purpose, or at least they can't cater simultaneously to AI's speed, depth, and novelty. Consider the following three points.

Executive education has limits

Take executive education from universities. One problem with it is that the teachers, and certainly the courses' delivery model, are not calibrated for promoting "multilingualism" which is essential in making technology happen, as it enables people to connect (a) the new technologies with their (b) domain expertise, as well as the (c) capability to design human-centered new processes that are delightful for employees and customers. Executive education struggles with delivering granularity around individual domains and students are asked to connect the dots, which often results in a big loss in translation and real-life impact. Many universities are also relatively slow in adapting to a quick change in technology's frontier: many teachers at top institutions love to do very future-oriented research, which is not necessarily implementable in the short term by the students, and is therefore partly irrelevant; others are too far from the current best practices being implemented in companies.

Online learning has limits

MOOCs like Coursera or Udacity could be an answer. They're cost-effective, they can be accessed from anywhere, and many courses are fresh. Udacity is now just being bought (at a discount compared to previous valuations) by IT services firm Accenture which intends to use it for its overall transformation capabilities and accelerate its - and its clients - AI skills. While brilliant at teaching many things, even the best are not fully fit for the new purpose in isolation, and it takes more work - namely embedding knowledge into the work, and heeding to its granularity - to get optimal results.

One more limit: your managers

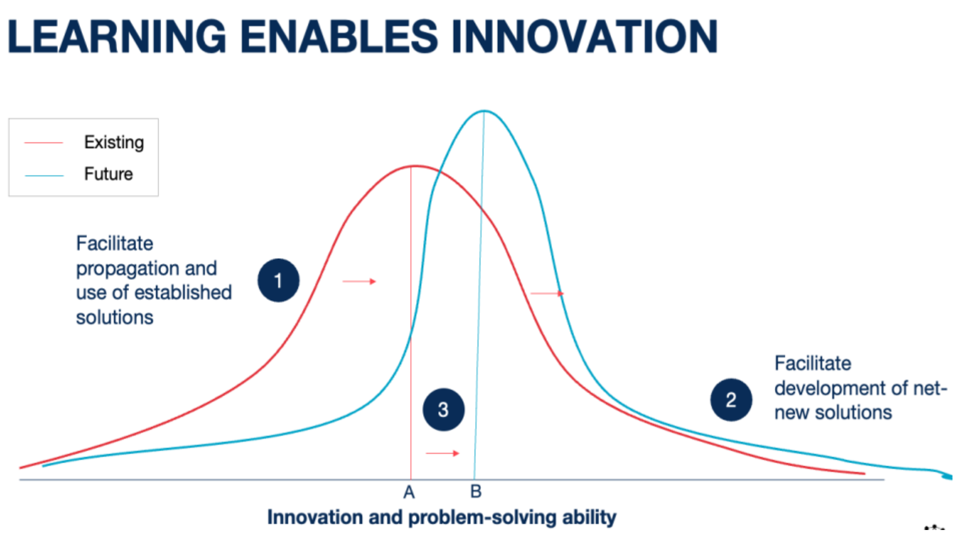
Corporate Learning & Development (L&D) is full of dedicated and smart people but it is also often structurally underpowered - not just because of its resources, but because much of the effort goes into convincing people to learn. One big reason is that too many managers are not considering learning (and teaching) part of their day job, outsourcing that to L&D departments which in return tends to become too much of a courseware procurement arm. Much of the expenditure ends up in executive education (see previous paragraph), as external courses are often more entertaining and add visible value to participants' CVs. Most strategically, despite the rhetoric, learning is too often considered a discretionary individual motion, an effort, and an investment, separate from innovation at scale and company vitality.

This, and much more, makes many skills architectures unfit for purpose in today's environment.

Beyond marginally incremental solutions: scoping the problem in a different way

Trying to force fit this problem into standard organizational solutions will result in suboptimal solutions. This AI wave is the most important tech shift of the last decades. Failing to recognize that will impair our path to deploying AI's power at scale.

As I wrote, the scope of this problem isn't just an HR one. This is not just beyond L&D, but it is also beyond a workforce planning problem. It is also an employee engagement and productivity problem, and ultimately, because of its connection with innovation (see chart below and this article), it is a *strategic capability* problem, with an attendant hard-dollars business case, that should sit firmly in discussions with the CEO and their management teams.



First - what should people learn?

We must give much clearer guidance to workers about which skills they would need to build, to get which jobs in the near and longer term.

The first part of the job is planning, starting with an elusively simple question: *what tasks are being performed in your company, and will need to be performed even more based on the company's* *growth*? That requires work because most companies track jobs, not tasks, and their prevalence.

Based on this information, one can approach the next question: *which tasks are impacted by today's and tomorrow's (predictive AND generative) AI?*

Against this backdrop, the third question is - *which skills should humans learn in the age of AI*? Certainly not the skills for which AI will be better than us - with some exceptions in case they're foundational to human development. So you're left with a broad palette of themes like logical and critical thinking, problem-solving, grounding and critiquing machines, collaboration with people and machines, ethics, and generally domain expertise to keep the machines' output in check while connecting the dots and asking the right questions. That's in addition to skills to design and build AI-enabled systems which, depending on the type of role, may or may not require technical skills (for instance, prompting and no-code software development are increasingly democratizing a range of tasks previously reserved to coders).

We also need to be practical. We should focus most resources on the skills that help people design and build things around what AI can do today and tomorrow ("horizon 1" = assistants) and help people be better, more scalable *humans in the loop*. Then begin focusing a smaller set of resources on AI for process transformation ("horizon 2"). Finally, allocate an even smaller quantity of resources on "horizon 3", focusing on what AI can do in a few years (such as organizational transformation through autonomous agents that completely reshape workflows). There is no one-size-fits-all: if you're an innovation-oriented trailblazer, the proportion of resources on horizon 3 may be higher than in a fast follower.

Second - what is learning here? The flow of work.

In answering this question, we are often biased by what learning *institutions* have been, which in turn was biased by their past and current delivery model (industrialized, standardized) and the participants' incentives (academic research, and market signaling of student quality). We are also swayed by assuming that learning is something that happens *outside* of work, either through school years or through executive education. MOOC set out to solve some of this, but it is incomplete in isolation.

The reality is that for **most working adults learning *already* happens every day in the flow of work, and that flow of work is where much additional learning should happen.** There are at least two reasons for this: first, because work's "gravitational pull" is enormous, and most adults (and their managers) struggle with prioritization of learning over daily tasks; and second, because adults learn by doing and reflecting on the result of their actions, and need contextualization and connections with what they already know and do (their domain expertise) to be able to retain and reuse the new knowledge.

This is not just desirable but also feasible and viable. Traditionally this vision has been hard to deliver. But today we have the technological tools and the methods to change the game - especially for technologically rich and fast-evolving topics like AI (more on this in a moment).

If we think of learning this way, then educational courses become only one part of the knowledge-absorption process, but there is a lot more to it that we can bring to bear thanks to new technology and practices. For example, *Generative AI can radically transform access to knowledge* by enabling conversations with a rich and engaging knowledge corpus, blurring the boundaries between knowledge management and learning. Think about reasoning-support engines like large language models (LLMs) when they're properly scaffolded (like the MIT Supermind Ideator, the OpenAI GPT's Apta my teams and I worked on); or about knowledge retrieval ones like Perplexity.ai or Stardog (and soon new Microsoft capabilities), and McKinsey's Lilli. Generative AI can also engage in a proper dialogue with the learner, particularly by asking useful questions - we are seeing that already in places (such as K-12 education Khanmigo by Khan Academy's Sal Khan).

After workers have learned the basics, they can apply them to their jobs which can deliver real-world improvement and innovation and yield organizational learning at the same time as they're likely captured in interactions with other colleagues and the creation of new artifacts. As they become more T-shaped (that is, able to master the basics of many fields adjacent to the one they're deep in), people can employ the new concepts and ideas directly into their work, coming up with new ideas that are often incremental applications of existing practices, some of which may be new for them and their companies, but at times can yield groundbreaking recombinations of concepts from different fields.

Building on successful efforts (e.g., the work done by my teams in large enterprise environments) AI can now turbocharge these processes, once more blurring the boundaries between learning, knowledge management, collaboration, and ideation.

All of these concepts can be used to deliver, at scale, a rich learning experience in the flow of work. They can *turn the flow of work itself into learning.*

How to deliver learning for AI, in AI's age

How does one deliver the ideas discussed so far? While we can't do justice to an AI learning architecture here, below are some of the most novel and distinctive elements for an AI-ready skills infrastructure.

Guide people to what they need to learn

First, back to workforce planning: **determine your current and future "skill stock"** and the gap compared to your skill need. That is - thinking of skills as components of a finished product (a role) whose demand is dictated by your company's business need. What skills do people need, as machines take over some tasks, and people will work with them to control them and extend their own capabilities?

There are three parts to this question: (a) what machines can do that humans do today (and in the future); (b) what humans can do to machines to make that happen; and (c) how to design processes that help machines and humans perform well. To answer, one needs to understand which skills are needed in individual roles and move away from a traditional taxonomy of jobs that is too coarse. And if this feels hard, remember that we have done this before, with other technologies.

That should be your target function.

Design and deliver the learning infrastructure to support the learning experience

Learning isn't an L&D process and is not just done by learners. Reimagine the process of learning as an experience for the personas involved, starting with a competence assessment to help learners understand their gap for their specific role (which leverages the workforce planning analysis from the previous point), and then starting absorbing knowledge by reading, viewing, interacting with others, and applying it in the flow of work. Other examples come from the direction that Microsoft and Salesforce have taken, enabling rich embedment of learning resources, such as those from external providers, into their environment. All of this has traditionally been hard to do at scale - but it is not impossible, and it doesn't take a prohibitive amount of resources, as demonstrated in my team's work with Genpact's Genome, for instance. Today we have both the technology (thanks to AI and other digital tools) and the practices (thanks to organizational science related to collective intelligence, including our work at MIT CCI) to harvest the collective intelligence of an organization as it identifies use cases and experiments in real-time, somewhere in your ecosystem. That will help connect theory with practice and enable implementation so that there's no "lost in translation" when workers absorb and apply their new knowledge.

That's why it is legitimate to say that learning, collaboration, and knowledge management *belong to one discipline*.

One way to do some of this is by **harvesting the "exhaust"** of communities of practice (say, on Slack or Teams) which can then turn into a complement to formal knowledge management. Over time, one can use AI to crystallize this knowledge and build "digital twins" of practice leaders' knowledge corpora, for example, by building on top of LLM's APIs (Genpact AI guru was an early example starting in 2023) or in the form of relatively simple, and very flexible, OpenAI GPTs.

Learning during the flow of work and experimentation should be strongly supported. For example, through safe AI sandboxes workers are provided contextual guidance and can experiment - while trying to solve real problems - and then discuss the results with their peers. For generative AI, that means improving things like how to prompt, logical tools for critique and triangulation, and at the same time the application of domain expertise. Building company-specific tools, for example, GPTs, can be helpful, as long as they're kept private when required.

Think about the entire delivery stack, and especially the data layer. Supporting all of the above, there is the traditional L&D stack, IT infrastructure, HR policies, etc. There is also a massive implication on data strategy - data for up/reskilling in the age of AI is as important as ever because it trains machines who train people. (Just possibly, Accenture's acquisition might have been predicated on the fact that Udacity has millions of test submissions enabling extensive data labeling). Both structured and unstructured data are important - for example, all the digital exhaust through Teams or Slack channels including, for example, curated recordings of important sessions.

Think big but start small and be ready to scale. Where would one begin, to deliver some results in the short term? I recommend starting with your leaders, and with a few people in specific business areas that will be impacted first by the new technology advances: for example, parts of customer experience, marketing, salesforce effectiveness, and R&D.

Some of the above could be done with an incisive and time-bound "*AI readiness Plan*" whose initial interventions could be delivered within a few weeks:

1. an **assessment** of individual competencies
2. a **mapping** of those against role-specific needs and resulting identification of the individual gaps
3. then learning the **foundations** for each level of competence, in

isolation, also with the help of a digital, AI tutor

4. followed by **interactions** with peers supervised by experts focused on using the knowledge on real problems.

5. then ongoing **coaching** happens over a much longer period In the flow of work - facilitated by networks of people including not just instructors, but also communities and indeed digital, AI coaches.

architectures can't do that without innovating themselves.

We have recently experienced a progressive shift from learning as a standalone motion, to learning in the flow of work. Because of AI and its impact, we will likely see another subtle but world-changing move from learning in the flow of work to realizing that **the flow of work *is* learning**.

Start by extending what you have, and building the new infrastructure on top of it, today.

SIMPLIFIED

	Skill stock and gap identification (individual and organization)	**Learning the foundation** (at each proficiency level)	**Interacting with others** (individually, cohorts, segments)	**Consolidating through work** (actual application)
When	Mostly early	Early, then ongoing for refresh	When foundation partially done, then ongoing	When foundation partially done, then ongoing
Content Knowledge	N/A (informs tests)	Courseware, content harvested from knowledge bases, external information feeders	Harvesting of knowledge from interactions (e.g. seminars, threads)	Harvesting of artifacts produced during work
People-network knowledge	Crowdsourced retrieval of experts in the network	Knowledge management harvests knowledge from experts	Interaction with skilled individuals, synchronously and asynchronously	Flow of work. Project marketplaces, peer-based assessments
Tech-supported knowledge	Quizzes/tests Skill taxonomies (by role) AI-enabled sensitization (e.g. after meetings, communications, deliverables etc.)	Conversational tutors Quizzes/tests AI-enabled tutors	Synchronous: seminars, peer reflections, small-group learning "watercoolers" Asynchronous: threads summaries, forums feeding AI corpus, AI coaches	Task-contextual retrieval of relevant knowledge and people "Teachable moments" AI assistants e.g. reasoning engines, ratings, critique

Beyond the Turing Trap

Skills are a strategic, necessary lever in delivering the world-changing innovation that the age of AI promises. They will help people gain the capabilities and confidence to help them benefit from the introduction of new technology. But skills

Beyond "Human in the Loop": Reliable AI in Enterprise Workflows

GenAI models are still too inaccurate and unreliable for many processes. But so are humans, in isolation. We know how to scale the use of labor input, and deliver quality output - why can't we do that for digital labor?

The world is grappling with identifying the *use cases* where generative AI can add value, and *how* to deliver that value. Most of the "how" discussion revolves around the improvement of the models, with much of it focused on how to improve the reliable accuracy of the responses, through methods such as Chain (and Tree) of Thought, self-consistency, retrieval augmented generation (RAG), and data quality processes among others.

In the absence of sufficient reliability, delivering many important use cases will be a steep climb - as witnessed in lots of current pilots. Some are already fearing we are nearing the Gartner Hype Cycle's "Trough of disillusionment".

However, we might be making our slog steeper than we need to, I argue in the rest of this essay. We must explore additional ways to "industrialize" the use of generative AI so the *business processes* where AI is embedded - and not just the algorithm - deliver the requisite accuracy. To do so, it is useful to take some inspiration from scientific process design and operations practices that we have used for decades.

Let's discuss this step by step.

We have never seen machines like these before

The current situation is one where new generative AI machines:

1. **behave in some important respects like humans**, which is a relatively new condition for traditionally deterministic production engineering.
2. **often fall into *many* of the typical categories** such as tool, assistant, peer, or manager (borrowing from MIT Prof. Malone's categorization).

That confuses many technology product managers, IT leaders, and solution architects.

Not coincidentally, AI researchers are acknowledging the crucial role of "**human in the loop**", which means using competent people to supervise machines.

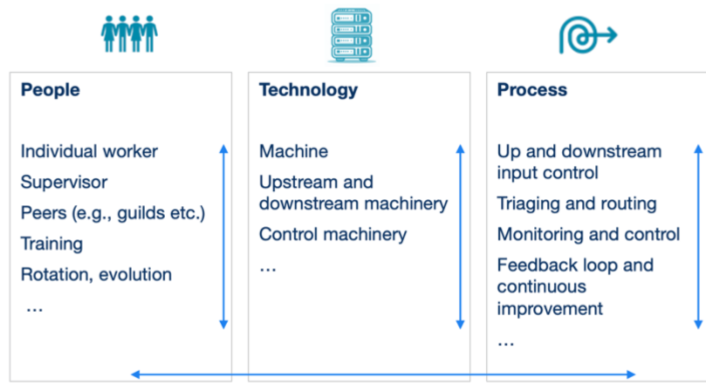
Now notice how the situation is similar in some ways to the role of *supervisors* in traditional work. The issue there was – and is – that supervisors don't scale, and they, too, are not error-proof. Supervision has evolved in recent times, as part of faster digitally-enabled business processes such as those in the front/middle/back offices of organizations. But AI can process things orders of magnitude faster than humans, so human supervisors would continue to be a bottleneck.

"Human in the loop", now that labor is generative AI-based

Or better, if rephrased: how can we build pervasive, scalable - industrialized, even - supervision of AI?

Granted, we are now talking about *digital* labor's performance. But as I argued elsewhere, organizational theories and practices from Taylor to Ford, Lean, TQM, and Six Sigma, have for more than a century improved the quality and reliability of *human* labor, and they're relevant now with digital labor. For one, the AI's precision vs reliability problem mentioned above is germane to, in operational terms, a defect-rate vs variance problem. **What can be ported from that scientific process design experience to help today's AI-enabled processes?**

The first portable concept is to that we **should focus on the entire operational delivery stack**, not just individual technologies, or people, **as the "unit of production"**. Delivering processes at scale requires designing the entire stack. The stack has evolved but it is still relevant. Much current AI research gives a short drift to the *people* part of it, beyond some nods to reskilling. It seems to also largely ignore the *process* side of the story, beyond the technical AI workflows e.g., AI Ops, a derivation of MLOps. (One exception is the recent HBR Article titled "How AI Fits into Lean Six Sigma", authored by Matthias Holweg, Thomas H. Davenport, and Ken Snyder.)

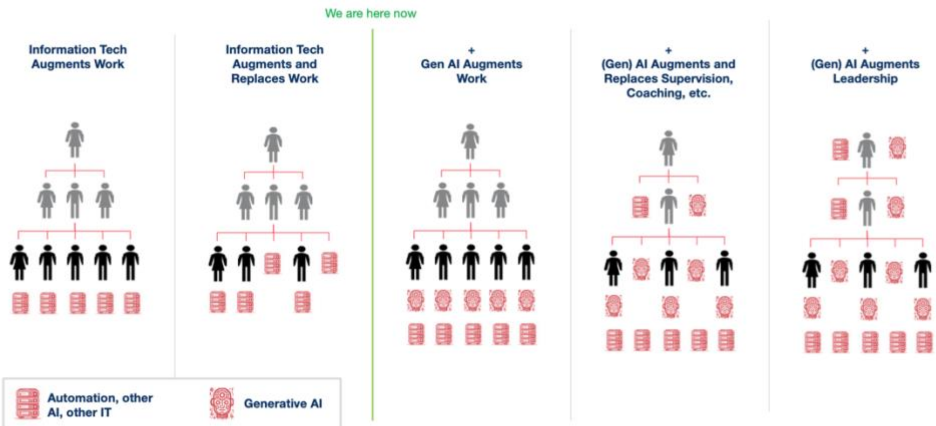


Let's also remind ourselves of how the shortcomings of individual production inputs (unreliable human labor or machines) have been holistically addressed across the operations delivery stack, and how that evolution could give us useful ideas, as summarized in the table.

SIMPLIFIED

	Pre-industrial process management (e.g., artisanal)	Scientific process management (Taylor/Ford/Lean / TQM/Six Sigma)	Scientific process management + digital / AI	Scientific process management + digital + Generative AI
People	Individual craft, apprenticeship	Specialization; training; feedback loops; collective involvement	"Design"/"build" through agile methods and new competencies; "run" through better user guidance and support provided by software	Data science competencies emerge; operators learn how to complement stochastic models
Process	Personal experience, small / slow processes	Explicit; tightly orchestrated; continuously improved; reduction of waste; customer value	Reimagination through superior information exchange, dynamic workflows, automated decisioning; human-centered design, CX	Human labor's tasks are unbundled and recomposed into new workflows; industrialize creative, knowledge tasks
Technology	Individual or animal mechanical power, or small machines / tools	Information technology (including non-IT e.g. Kanban cards, Andon cord); tight integration of production machinery	Systems of records, systems of engagement (e.g., workflows); RPA; AI (ML)	Generative AI complements other automation/AI technologies. Less deterministic, more probabilistic

In the diagram below, we build on these ideas to imagine organizations that increasingly accommodate the use of comparatively less reliable generative machines in addition to the others we employ, alongside frontline operators, supervisors, and leaders.



Therein lies the main intuition.

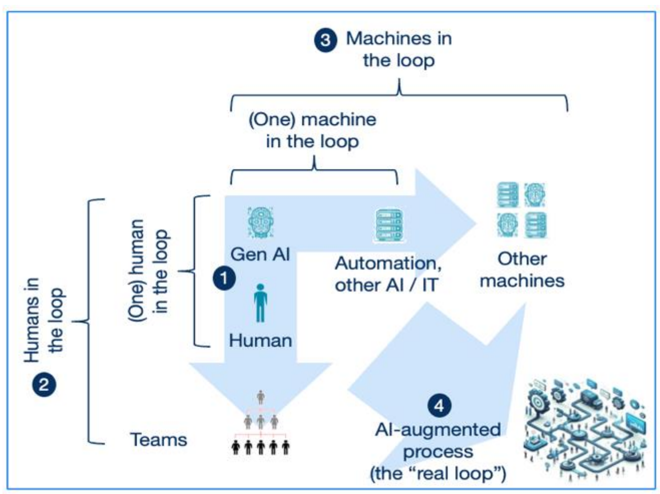
What if we didn't consider Gen AI machines as standard machines?

As a somewhat crude but clear example, think of our new machines as the equivalent of millions of interns. Resourceful and energetic, imprecise and somewhat unpredictable. And cheaper and much more scalable than real ones. How would we design organizations, processes, and other technologies around them to harness their power?

What if we could improve the new machines' performance (in particular, accuracy) by learning from what we traditionally do with humans in operational settings? This means for instance leaning hard on controlling their quality through the collective intelligence of other workers (supervisors, peers, etc.), as well as of other types of machines, and certainly relying on specially-designed processes. Supervision, in other words, *doesn't need to be individual, human, or hierarchical*.

Think of it like this: **a new generation of AI-augmented processes, designed to harness the power of machines and people, is *the loop***.

Some initial ideas resulting from this entire thought process (scientific process management organizational theories and practices, collective-intelligence methods) are shared below. They are not mutually exclusive, and certainly not collectively



exhaustive. They also require nuance or possibly downright overhaul in specific industries or processes. But they can provide an immediate checklist to the business architecture teams on the ground. At the very least, they can inspire an exploration beyond the current media-grabbing, but hard to operationalize, headlines.

Compared to that world, let's highlight four differences upfront.

Analyze like it is 2024. First, data science has firmly arrived in the process world. Tools like process mining, especially in combination with machine learning and other advanced data techniques, provide an unprecedented opportunity to discover issues both in individual activities and in the end-to-end flow of processes.

Build to evolve, not to last. Second, the frontier of technology's art of the possible changes with time. The upshot? New processes should be built for evolution, not built to last. This also means that the analysis phase in itself should be performed in an agile way and last weeks not many months, to avoid that its results are obsolete on arrival.

Design around the human too. Third, some of these conceptual tools were created before disciplines like human-centered design matured, which means that they (and their trained professionals) may not take adequately into account the opportunities and needs presented by the people side of the equation. Mitigate this risk by including service design experts in the program.

Beware of incrementalism. Fourth, many existing practitioners experienced in these tools have traditionally worked on incremental improvements, which means that they might have a bias towards anchoring themselves on the current processes and starting from there. In this time and age, radical reimagination is possible and can be achieved if the design team is given a chance. For instance, form a "red team" that looks at the process with a complete reimagination lens, while the rest of the team uses more traditional tools and approaches, and get the two to respectfully but firmly challenge each other.

This said many traditionally-used tools can yield value now in breaking down the process and prioritizing the right tasks by scoring them across importance and feasibility. The following is an initial list.

A new (old) lens that you can apply today

There are at least eight areas to explore:

1. Process *step* redesign around the human-AI frontier
2. AI System Development and Maintenance for Operationalization
3. Data Management and Quality Checks for Operationalization
4. Human-AI Interaction and Oversight

5. Performance Monitoring and Feedback
6. Operational Processes and Procedures
7. Triage and Case Management
8. Institutionalization of learning across these areas, for both humans and machines

In each, many activities and practices leveraging classic principles, described in the following table, are detailed in the appendix to facilitate their implementation.

Area	Activities/Practices
1. Process Step Redesign Around the Human-AI Frontier	a. Task Analysis b. Ergonomic Assessment c. Time and Motion Studies d. Value Stream Mapping e. DMAIC f. Theory of Constraints g. Elimination of Waste h. FMEA for Risk Assessment i. Benchmarking and Best Practices
2. AI System Development and Maintenance for Operationalization	a. Regular Maintenance of AI Systems b. AI Systems Risk Assessment c. Predictive Maintenance for AI Systems d. Continuous Improvement of AI Systems e. AI Systems Simulation and Modeling f. Flexible AI Process Design
3. Data Management and Quality Checks	a. Upstream Quality Checks in AI Data b. Supplier Integration in AI Data c. Back-feeding data from downstream operations
4. Human-AI Interaction and Oversight	a. User-Friendly AI Interfaces b. Cross-checking AI Outputs c. Buffer Zones in AI Decision-Making d. Error-Proofing in AI Interactions e. Optimized AI Monitoring Work Environment f. Standardized AI System Handovers g. Fatigue Management for AI-facing Teams
5. Performance Monitoring and Feedback	a. Training and Education b. AI-enabled Process Performance Monitoring c. Communication About AI Performance d. Downstream Quality Audits for AI Outputs e. Customer Feedback in AI Development
6. Operational Processes and Procedures	a. SOPs for AI Deployment and Monitoring b. Checklists in AI Operations c. Automation Aids in AI Supervision d. Quality Circles in AI Oversight e. AI Anomalies Reporting Systems
7. Triage and Case Management	a. Intelligent Routing Systems b. AI Performance Monitoring with Triage c. Dynamic Case Assignment d. Redundancy in AI Outputs e. Feedback Loops in AI Design
8. Strong learning loops for all of these areas	a. Systematic "data-to-insight-action" loop for learning b. Contribution of both people and machine networks c. Enlisting the power of technology to mine those insights d. Deliberate mapping of skills e. Institutionalize effective collaboration flows

Ignore these practices at your peril

The combination of people, process, and technology has always been a form of "collective intelligence" architecture – but was often a deterministic and hierarchical one. With the new class of AI machines, more of the past organizational and process models might become unwanted legacies. My view is that, at least for now, machines alone won't be good enough to substitute human labor in today's processes.

This feels in some respect similar to the first wave of smart automation of the mid-2010s - those who addressed the challenge as not-just-technology prevailed. This time, the urge to look beyond technology paradigms is even stronger - because the new machines themselves don't behave like the machines we are used to. Being able to harness the power of the new capabilities by designing our organizational and business **processes** as well as evolving the related people practices is likely to be **as important as working on the technology side of AI**.

To some, this may feel like less-novel work. But it is worth remembering that most business-to-business technology deployed in the world of work isn't really "plug and play", and architecting the *combination* of people/process/tech – as well as *transitioning* away from legacy operations, as we saw years ago in the first wave of enterprise AI and related automation - remains the key to delivering real-world outcomes.

Appendix

1. Process Step Redesign Around the Human-AI Frontier

Scientific process management's "ground zero" was to re-design processes so that each task - the "what" - could be performed better by a "labor input" - the "how/who". Traditional workflows would be disassembled into more elemental components, and those components recombined into a new workflow so that each step could be performed by the most appropriate resource (available people and/or machines) - and hence with fewer defects.

a. Task Analysis: A detailed analysis of tasks considering human capabilities and limitations. It helps in identifying tasks that require human traits like judgment, collaboration, and some types of problem-solving that AI currently cannot replicate.

b. Ergonomic Assessment: Evaluating tasks for physical and cognitive ergonomic principles to decide if they should be human-operated or automated. This is particularly important as machines become truly multimodal.

c. Time and Motion Studies: Analyzing the time and motions involved in each task (which can also be done through multimodal AI) to determine if AI can perform the task more efficiently than humans.

d. Value Stream Mapping: This tool helps in identifying all the steps in a process and categorizing them into value-adding and non-value-adding. AI can be deployed to automate non-value-adding tasks that are routine and repetitive.

e. DMAIC (Define, Measure, Analyze, Improve, Control): This framework can be used to analyze processes and identify steps where AI implementation can reduce variability and improve quality. Tools like control charts can monitor process performance and determine where human intervention is more effective compared to AI.

f. Theory of Constraints (TOC): Using TOC to identify bottlenecks in a process and determining whether AI can be used to alleviate these bottlenecks more effectively than human labor.

g. Elimination of Waste: AI can be utilized to eliminate waste in processes, such as waiting times, over-processing, and defects.

h. FMEA (Failure Mode and Effects Analysis) for Risk Assessment: Analyzing the potential failures in each step of a process and the effects of those failures to decide whether a task is more suited for AI or human intervention.

i. Benchmarking and Best Practices through Comparative Analysis: Looking at how similar processes are managed in other organizations or industries, and the role of AI in these processes. It is likely that some industries, such as financial and professional services, will take the lead in deploying generative AI at scale, and will show what types of tasks are best suited, and how to engineer the respective process.

2. AI System Development and Maintenance for Operationalization

This area, like #3 below, is closer to the bulk of current AI research, but more tightly coupled with the upstream and downstream operations processes, which means for instance sharing clear KPIs across the chain, and employing cross-functional resources (including "T-shaped" people who understand each other's domain).

a. Regular Maintenance of AI Systems: Regularly update and calibrate AI systems, including updating models with new data and fine-tuning parameters (see next section too).

b. AI Systems Risk Assessment: Conduct risk assessments for AI systems to identify potential sources of errors or biases and develop strategies to mitigate these risks. Involve operational resources (such as process operators) in conducting these tests.

c. Predictive Maintenance for AI Systems: Use predictive analytics to anticipate and address potential AI system issues, monitoring performance indicators to predict model degradation. Inform predictive analytics with continuous feedback loops from the actual operations (see next point).

d. Continuous Improvement of AI Systems: Establish a culture focused on continuous AI improvement, regularly reviewing and updating AI models and algorithms based on performance data.

e. AI Systems Simulation and Modeling: Use simulation tools and practices to anticipate AI behavior under different scenarios, identifying potential failure points or performance issues for preemptive adjustments. Include operational resources, such as process operators, in those stress tests.

f. Flexible AI Process Design: Build flexibility into AI systems to adapt to varying conditions and inputs, using models capable of handling diverse input types and adaptable systems. Design for seamless handoffs to human and other technology resources (different tools), so that AI-based processes can “fail gracefully”.

3. Data Management and Quality Checks

a. Upstream Quality Checks in AI Data: Implement early-stage quality checks in data processing for AI systems, ensuring data accuracy and completeness in the most important areas (apply 80/20 Pareto analysis when needed, with the help of actual operational domain knowledge pinpointing the most critical areas).

b. Supplier Integration in AI Data: Collaborate closely with (external and internal) data providers for AI systems, ensuring high-quality and relevant data for training and running AI models. Leverage these practices for the rest of your operational data production (see next point).

c. Back-feeding data from downstream operations: The output from models will be continuously verified in production, and the feedback from that can be fed back into the model in the form of fine-tuning, Retrieval Augmented Generation, knowledge graphs improvement, etc.

4. Human-AI Interaction and Oversight

This section is partially derived from existing human-computer interaction (HCI) practices but with a specific focus on operationalization at scale, as well as leveraging the collective intelligence of organizations.

a. User-Friendly (Human-Centered Design of) AI Interfaces: Design AI interfaces and protocols that are intuitive for human supervisors, with user-friendly dashboards and clear alert systems. Design those for the kind of supervisors that exist in operations and their varied capability levels. Ensure that the human is kept alert by appropriately sequencing the type of input provided by the AI (e.g., asking for the human's input first, or asking for a critique of the AI's output).

b. Cross-checking AI Outputs: implement cross-checks for AI outputs, either by another AI system or manual human review, especially in high-stakes scenarios. Enable retrieval of supporting evidence from existing fact bases such as standard operating procedures. Ensure the AI provides justification of the logic followed to arrive at its conclusion and use different types of AI (and other technology) to triangulate and verify the consistency of the output, to identify the most likely choice, or at least estimate the degree of confidence. If needed, design for human-network reviews, or at the very least with clear escalation paths to more competent (and expensive) resources, whether machine or human.

c. Buffer Zones in AI Decision-Making: Create intentional buffer zones in AI-enabled processes for increased human oversight in critical areas, involving additional manual reviews or secondary AI or other technologies' checks.

d. Error-Proofing in AI Interactions: Design AI systems with built-in safeguards to prevent misuse or misinterpretation, incorporating constraints and user interface design (such as in-built triangulation, and spot supervisor checks) to avoid errors.

e. Optimized AI Monitoring Work Environment: Optimize the work environment for AI monitoring teams, ensuring they have the necessary tools and a conducive environment for detailed analysis. This means instrumenting the entire business process to gather upstream and downstream data that can improve performance analysis and explainability of root causes.

f. Standardized AI System Handovers: Establish standardized procedures for AI system transitions, such as team changes, model updates, or phase transitions, ensuring consistency. For instance, models' prompts and their APIs may behave slightly differently after the release of new versions.

g. Fatigue Management for AI-facing Teams: Implement strategies to prevent overwork and cognitive overload in AI monitoring teams, such as workload management, adequate breaks, and workflow sequencing changes. In particular, ensure that humans don't over-rely on machines, and "fall asleep at the steering wheel".

5. Performance Monitoring and Feedback

a. Training and Education (for AI-facing Teams): Continuously educate and train AI teams on advancements, ethical considerations, and best practices in AI deployment and monitoring. Similarly, train the human operators and supervisors (for instance, with specific critical-reasoning skills) so that they can detect early signs of misalignment.

b. AI-enabled Process Performance Monitoring and Feedback: Monitor process performance metrics continuously and collect feedback for refining AI models and algorithms, as well as improving the interfaces with processes and people.

c. Communication About AI Performance: Maintain clear communication channels for discussing AI performance, updates, and issues, including standardized reporting of AI metrics. Also include synchronous and asynchronous forums for operators and their supervisors to share ideas, practices, and concerns.

d. Downstream Quality Audits for AI Outputs: Conduct thorough quality audits at various stages of AI output generation, including random sampling and full-scale reviews of AI-generated content. Do that by also using process data upstream and downstream, to improve root cause analysis. Make “AI-enabled process-forensics design” an important discipline, staffed by cross-domain professionals.

e. Customer Feedback in AI Development: Incorporate end-user feedback into AI development and refinement, aligning AI outputs with user needs and expectations, and identifying improvement areas.

6. Operational Processes and Procedures

a. SOPs for AI Deployment and Monitoring: Develop and adhere to SOPs for deploying (including operational processes use, not just IT production), monitoring, and updating AI systems, including guidelines for AI failures and unexpected behaviors.

b. Checklists in AI Operations: Develop general checklists for AI deployment (same definition as above), maintenance, and troubleshooting, ensuring consistent processes are followed, and involve AI-facing teams in optimizing and localizing those checklists.

c. Automation Aids in AI Supervision: Implement automation and technological tools to assist in monitoring AI systems, such as anomaly detection systems for unusual AI behavior. That also considers the differential capability and capacity of front-line staff and their supervisors, as well as staff in the downstream process.

d. Quality Circles in AI Oversight: Foster a collaborative team environment for discussing AI performance and improvements, encouraging collective problem-solving. Ensure that everyone can contribute, not just the IT experts or the supervisors, by creating diverse forums. Enlist the help of generative AI to orchestrate the feedback collection as well as its summarization.

e. AI Anomalies Reporting Systems: Establish a system for reporting (not just monitoring) AI anomalies and errors, enabling detailed data collection for (technology, process, and people-intervention) debugging and improvements.

7. Triage and Case Management

a. Intelligent Routing Systems: Implement systems that analyze incoming cases and direct them to the most suitable AI tool or human operator (or their supervisor, or other machines) based on predefined (and possibly dynamically adjusted) criteria. Use the learnings from the areas above to adapt them over time.

b. AI Performance Monitoring with Triage: Monitor AI tools' performance (see previous sections for details) and integrate a mechanism to redirect cases if an AI tool suddenly or systematically underperforms or faces unsuitable cases.

c. Dynamic Case Assignment: Use AI systems capable of assessing their confidence in handling a case, redirecting it if confidence is low.

d. Redundancy in AI Outputs: Use redundant systems or parallel AI models for critical tasks to cross-validate and correct AI outputs, in an approach similar to the self-consistency triangulation currently attempted on AI models themselves.

e. Feedback Loops in AI Design: Integrate into triaging specific feedback loops for learning from outputs and human interventions, adjusting algorithms based on performance feedback.

8. Strong learning loops for all these areas

a. Systematic "data-to-insight-action" loop for learning: Ensure systemic learning for all of these areas by (a) designing and (b) implementing deliberate feedback loops and learning cycles across insight tagging, learnings' storage and retrieval, as well as intentional learning processes for humans and machines.

b. Contribution of both people and machine *networks:* Insights that lead to continuous organizational (machine+human) learning will be fragmented across the organization. The right forums and incentives need to be put in place to allow the effective and efficient flow of insights.

c. Enlisting the power of technology to mine those insights: Tagging incidents can lead to ML discoveries later on. Make sure that your teams don't waste opportunities to tag anomalies and their resolution, even if they're not able to do justice to them immediately.

d. Deliberate mapping of skills: People (and machines) with the right specialized skills are important to solve pointed problems. Those skills are often not easy to find, especially in large, decentralized organizations, and when those skills develop fast. Strive to document the formation of those skills and their association with individuals, so they can be retrieved and harnessed later.

e. Institutionalize effective collaboration flows that emerge, and proactively retire others: As the organization becomes able to come together and solve specific problems, observe the emergence of collaboration structures, especially the informal ones. Try to facilitate them, but ruthlessly prune anything that has become just an exercise of bureaucracy.

What is Leadership in the Age of AI?

The advent of artificial intelligence is reshaping the landscape of leadership. Some things don't change (the Why of leadership), but others (the How) might, at least partially. For one, we know we need leadership now - the magnitude of inflection is as large as any we have seen.

Here, I explore perspectives on what it takes to be a great leader in the age of AI, the surprising shifts in decision-making processes, and the emerging hybrid models that blend human judgment with machine efficiency.

In my view, the federating concept in all of this is that leadership has always been about managing, for all practical intent and purpose, somewhat autonomous systems through "high-leverage points." Unlike management, which often speaks the language of process and psychology, leadership speaks the language of systems dynamics, politics, and social psychology. CEOs don't have a button to press to lead companies; they influence them through leverage points such as incentives, culture, and strategy. How does that change now that these companies will be increasingly bionic, with humans and machines collaborating and influencing each other in networks?

Here's a short rundown.

The New Qualities of Leadership: Experimentation and Exploration

Experimentation and Risk-Taking matter.

One clear insight from our discussion is the premium placed on **experimentation**. Leaders today are expected to:

Take calculated risks: Experimenting with new ideas is no longer a long-term gamble that spans years—it's a rapid-cycle process where quick iterations can reveal what works and what doesn't.

Redefine risk: In the AI era, "risk" means embracing uncertainty with the understanding that failure is a fast and valuable feedback mechanism. Instead of waiting years and spending vast sums to test an idea, AI can accelerate the cycle of learning.

Exploration: Asking the Right Questions

The ability to **explore uncharted territories** is crucial:

Ask questions instead of just providing answers: With AI platforms like ChatGPT readily available to generate answers, the true differentiator is the leader who can ask insightful, probing questions that spark innovation.

Pioneering partnerships: True leadership means venturing into areas no one else has and collaborating with unexpected partners. This form of exploration leverages AI's capability to process and learn from vast datasets, while human creativity maps out the unexplored.

While technology provides answers at scale, it is the human talent for asking the right questions and connecting disparate dots that will set tomorrow's leaders apart.

The Hybrid Decision-Making Model: Human-AI Collaboration

Beyond the Binary: Reject Extremes.

The conversation highlights a crucial prediction: rather than a full handover to AI (the "werewolf model") or complete human autonomy, the future will be **hybrid**:

Cyborg Decision Making: The optimal approach is a balanced model where AI handles routine, data-intensive tasks, and human leaders focus on creativity, strategy, and ethical judgment.

Maintaining human involvement: Leaders must avoid the trap of letting technology override human insight. Just as pilots must keep their skills sharp despite advanced autopilot systems, leaders must "keep their hands dirty" by engaging directly with customers, teams, and the ever-changing business landscape.

Building System Architectures for Hybrid Decision Making

Effective leadership in this new era involves designing organizational systems that:

Incorporate both AI insights and human intuition.

Develop infrastructures for continuous learning, collaboration, and knowledge management.

Foster agile cultures where decision making is distributed across multiple levels and informed by both data-driven insights and human judgment.

The real competitive advantage won't come from having access to the most powerful algorithms—it will be determined by how well leaders can integrate those algorithms into a system that amplifies human ingenuity.

Four Leverage Points for AI-Enhanced Organizations.

Several leaders emphasized a framework where effective leadership harnesses AI to impact four critical areas:

1. Network Topology

Mapping skills and resources: Understand where the talents lie—both human and machine. That's where the intelligence is, and what influences the behavior of the system.

2. Learning Infrastructures for Long-Term Thinking

System Two Thinking: Develop processes that enable teams to tackle uncertainty and plan for long-term challenges, mirroring the cognitive shift from fast, reactive decisions to more deliberate, reflective ones.

Amplifying learning: Use AI to continuously assess and optimize the distribution of skills across the organization.

3. Enablement and Knowledge Management

Beyond traditional methods: Leverage AI to capture, manage, and disseminate knowledge more effectively than ever before.

Learning organizations: Harvest teachable moments from daily operations, which are integrated into the daily flow of work, ensuring that lessons are retained and applied.

4. Collaboration Infrastructures

Building the "synaptic" structure: Just as the brain relies on complex synaptic connections for higher function, organizations need robust collaboration tools, practices, and culture - among others - that foster innovation across all levels.

The inimitability of a system that blends AI with human networks may well become the ultimate source of sustainable competitive advantage—one that is far harder for competitors to replicate than any single technology or algorithm.

The Evolving Role of Skills: Beyond STEM

While technical skills in math, computer science, and algorithm development remain essential, the conversation reveals an equally critical, and sometimes overlooked, set of competencies:

The Art of Questioning: In a world where algorithmic answers are readily available, the power lies in asking the right questions.

Interpersonal Relationships and Trust: As AI commoditizes technical knowledge, the human capacity to build trust and foster meaningful relationships will become a unique differentiator.

Holistic Problem Solving: Leaders must connect the dots across disciplines, using insights from social sciences, history, and human behavior to guide strategy and decision making.

As technology levels the playing field on many technical fronts, the unique human elements—creativity, ethical judgment, and interpersonal trust—will rise in prominence, making them indispensable leadership qualities.

Overcoming Pitfalls: Function vs. Competitive Advantage

Leaders must navigate several potential pitfalls as they integrate AI into their organizations:

Temporary vs. sustainable gains: While early adopters may see short-term benefits from using AI, these advantages can quickly vanish if competitors gain access to the same technology.

The real advantage: It lies not in the technology itself, but in creating a system that is uniquely inimitable. This might include: **Complexity and path dependency:** Systems built over time, with interwoven processes and cultural elements, are hard to replicate. **Unique organizational culture:** A strong, well-defined culture is perhaps the most elusive yet powerful asset a company can cultivate.

The greatest challenge is not mastering the AI tool but designing an ecosystem that leverages that tool in a way that competitors cannot easily duplicate. The focus must be on integrating human systems and technology to create long-term, sustainable value.

Conclusion: A Call to Adaptive, Hybrid Leadership

As AI becomes as ubiquitous as a calculator or a spreadsheet, the true competitive edge will come from our ability to use it thoughtfully, integrate it seamlessly with human judgment, and nurture the relationships and culture that technology alone cannot replicate. This is about management and organization design as much as technology.

In the age of AI, leadership is not about surrendering to technology but about harnessing it to amplify the best of what humans can do. The new leader is an experimenter, an explorer, and a system architect—someone who asks the right questions, builds resilient hybrid decision-making frameworks, and creates inimitable systems of innovation and collaboration. And one last thing: leaders, ask good questions of your organizations, and its AI. That's one thing that, for now, humans do much better than machines.

Why Some Quit, And Some Stay: A Surprising Take

Originally published: April 27, 2022, Revised May 2025

The engagement gap, the new implicit contract, and YOLO all affect the "Great Resignation." But here's a different perspective that leaders should urgently consider—one with some very practical implications for the retention of senior professionals.

For some time, some companies have been using enterprise social network analysis to identify the reason for employee attrition. The foundational social-network science on this is established (MIT's Honest Signals and Social Physics, for instance). We know that if we want, we can often anticipate and possibly prevent people's resignations by looking at their network signature. (Examples of the research here and here.)

But is there a more proactive picture? Is there a "prevention-medicine" equivalent, instead of just curing the individual problem when it happens? There's ground to think so.

Two reasons why good people go

There are many reasons why people leave but many of them relate to these two questions:

1. Do they **feel they're part of a (good) group**? Do their strong (e.g. manager) and weak (colleagues met at the watercooler) network ties give them a sense of emotional support, group "flow", as well as functional help?
2. Are they able to **make an impact** - a recognized and rewarded one - where they are? People's impact depends on which depends on being associated with the right things to do, getting access to knowledge and learning, doing work frictionlessly, and getting access to the right people to influence actions. The last one is especially important for senior people who typically work across organizational silos, and outside of fixed, rule-based workflows.

Interestingly, there is evidence that the answer to both questions depends on network structure. However, let's focus on the "ability to make an impact" because that's where many of the surprises are.

Having **impact through a network** means that, at parity of individual capabilities, (a) you can access hard-to-find (e.g. undocumented or unindexed) information and you can do that because your direct and indirect networks have it and (b) armed with the resulting recommendations, you can influence how things are done by nudging directly and indirectly the right stakeholders at the right time. Specifically, getting (b) done in the case of unpopular or counterintuitive decisions is what makes companies strong, and managers successful.

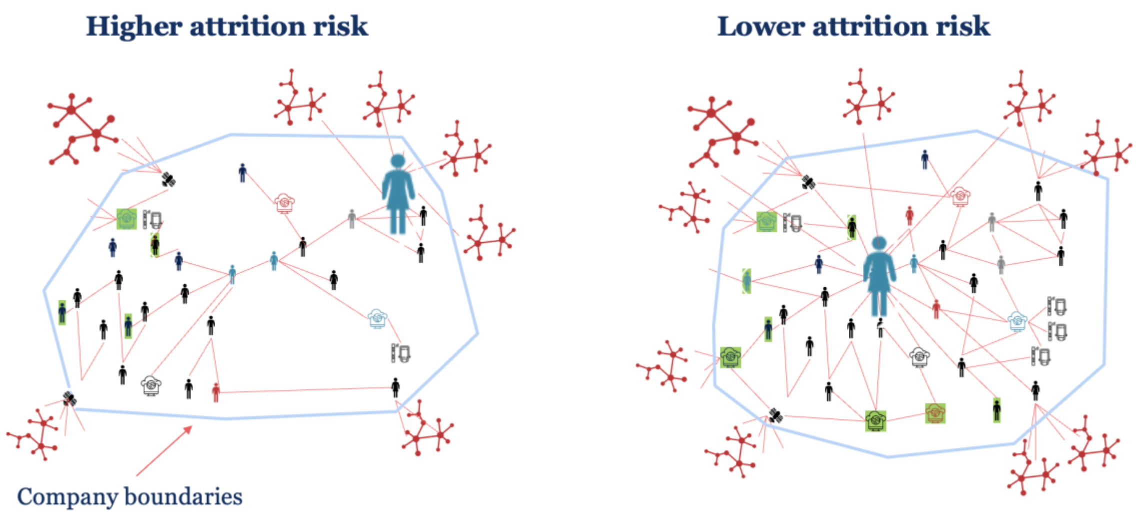
(The spreading of useful but against-the-grain knowledge has been researched thoroughly: some call it "complex contagion": it is information that, unlike easy "memes", isn't believed and amplified by the network unless the recipient hears it from multiple, unrelated sources. See here for instance).

The tipping point

Good people who fail to create a strong network in their current company often can't make enough impact, and consequently, they may have less to lose and could end up leaving. For them, *the skills that they have - not the network and the related social capital - are the most significant asset*. Those skills are marketable externally, and the loss of a company network only penalizes them a little.

Think, for instance, about people whose networks are particularly strong outside of the company (e.g., some salespeople, senior consultants): they would also have less to lose when changing employers. Also, research using LinkedIn data over five years, completed in 2022, shows that a large number of weak ties can help people find new jobs. That's especially true when remote work complements physical-office work, which makes moving elsewhere easier.

Contrast this situation with that of good people who have been in a company for long: their network's ability to generate impact represents a higher proportion of



their professional assets. And it's not as portable: that is, they would lose a lot of that if they moved elsewhere.

The simplified chart below illustrates this. The hypothetical at-risk employee is the blue, larger icon in the left quadrant.

The real brain in the brain drain

Step back for a second because there's an even bigger picture. We are saying that people use a "collective brain" to sense, remember, create, decide, and learn. That's what generates the real impact.

It is a form of what some call a supermind: a cognitive engine constituted partially by an individual's brain but also by the "neural" network structure in their organization or ecosystem.

That brain is marketable internally, or externally - depending on where the network is comparatively strongest.

What can be done

This has a clear implication: in companies where knowledge is easily retrievable and networks are readily accessible, relatively new employees typically become impactful more quickly, which should reduce the likelihood of their attrition.

Conversely, companies where the network is all-important yet tribal and knowledge isn't easy to retrieve will have an easier time retaining people who have been there for a long time: these employees' success is predicated on painstakingly building that network, which has now become the primary asset, and possibly a competitive moat against other employees.

So while the creation of internal networks is important for long-term retention, it can become a hurdle for others, especially but not exclusively new people, and that will make them attrite. But if companies **invest in enabling people's "network-based impact"** and improve **knowledge access** (including learning and knowledge management), the **collaboration tools** (e.g. virtual whiteboards, asynchronous conversations), and the ability to **network effectively** (including serendipitous encounters, affinity groups, communities of practice), they might be able to retain more of their best employees. Individual managers are responsible to make some of this happen, and they may need specific sensitization and training. But CEOs need to invest in augmenting the digital infrastructure that caters to the broader system and making the related change management happen.

This essay was written before the advent of Generative and Agentic AI. It is amply clear that new forms of AI can lead to amplified capabilities across these levers.

There's a clear choice to be made, right there.

Is Your Organization Intelligent?

The world is changing faster and becoming more unpredictable. Rigid workflows are often not fit for purpose anymore - they're becoming the equivalent of dinosaurs' brains: not helpful when the ecosystem is upended. Despite the massive shift due to the pandemic, many large organizations still struggle to evolve in terms of how they *collectively* sense, create, decide, act, and learn.

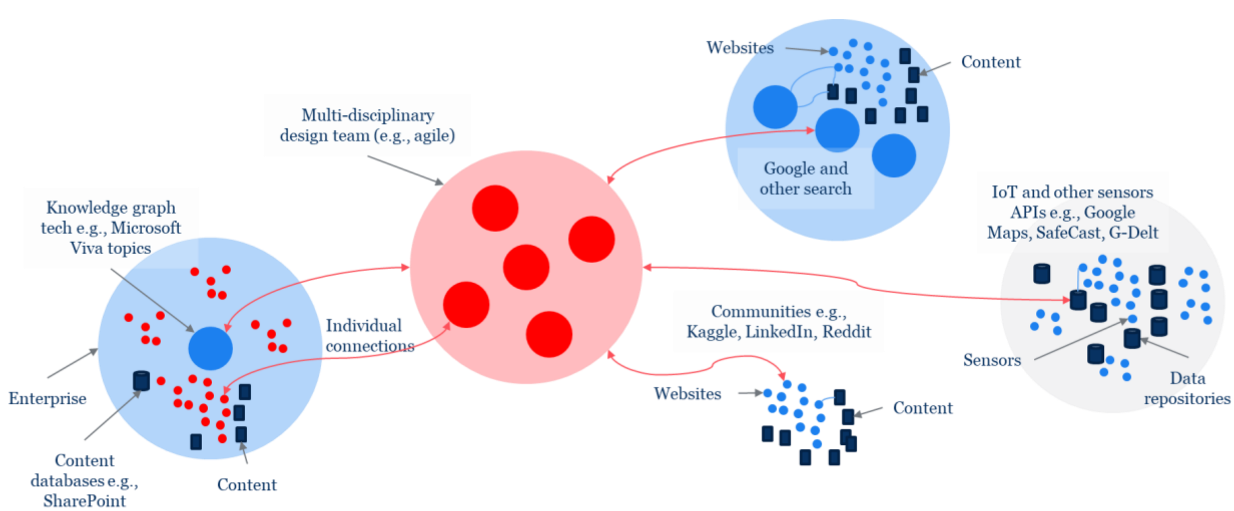
What design of that collective organizational brain provides a superior ability to adapt and compete? And can your teams put it into practice?

The graph of knowledge

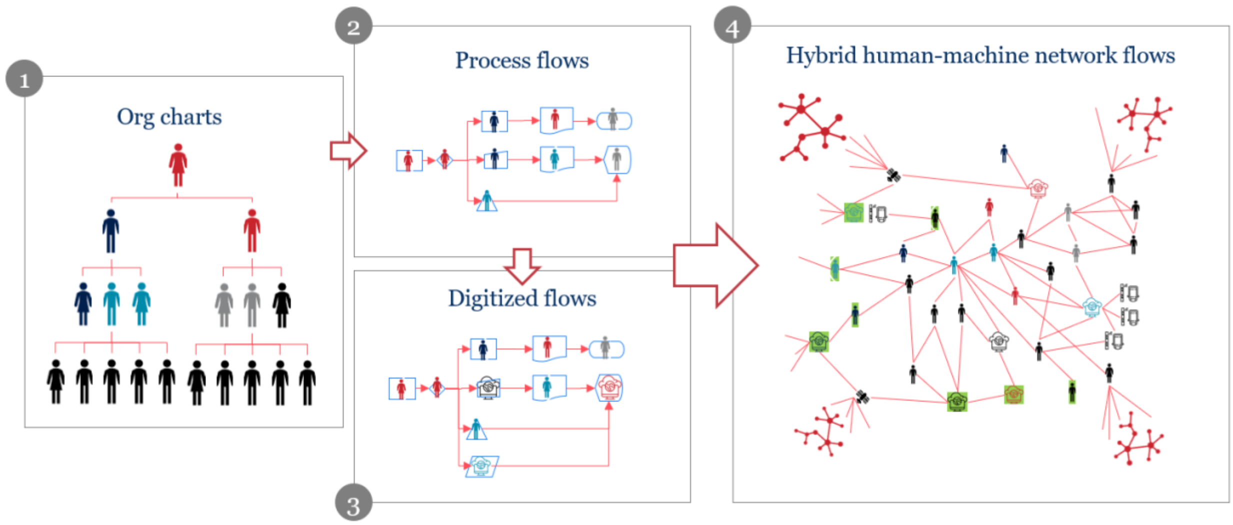
Today's organizations witness an explosion of knowledge available, absorbed, and filtered by organizational networks now solidly wired through Outlook and Slack. That knowledge is then amplified and evolved by internal and external social networks and meshed with continuous streams of other ideas curated by AI-based algorithms. These structures were in their infancy only five years ago, but they're increasingly the way business happens.

Not coincidentally, one of the fastest-growing data science spaces is knowledge graphs. Their biggest advantage is that they can document and process relationships, like what Google does with the world's knowledge.

What do those networks of knowledge look like? Look at the next chart, starting with the red "cell" in the middle, and think of an example that many executives experience daily: innovation teams (and people, the red dots) working on the next big thing or simply on business-as-usual continuous improvement. Their network structure looks like something like the following – and will increasingly add blue nodes, that is, networked machines that



complement and amplify the knowledge of people.



Most managers don't consider this picture when they design their organizations. This is because we don't learn how to design collective intelligence in school or as we rise through the ranks. But today, we can do much more than design org charts, workflows, and instill traditional management practices.

Beyond dated organizational design practices

How do we add to the traditional management methods (in Figures 1, 2, and 3 below) the ability to deliberately orchestrate networks across organizational boundaries (figure 4)? How do we architect such a system?

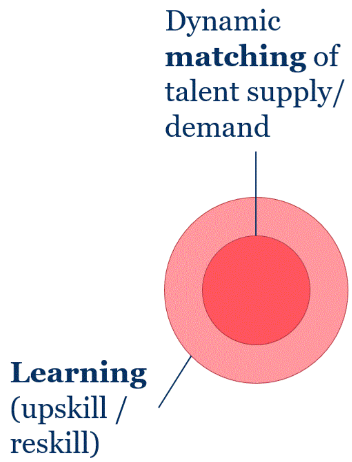
First, we need to unlearn. For instance, we likely need to abolish the conventional boundaries between disciplines—HR, IT, knowledge management, etc. The intelligence fabric depends on *blending* those practices.

Think of a salesperson using specialized documentation for a sales pitch created by a subject matter expert or a contact center agent prompted to respond to a client in a certain way based on a machine-learning algorithm that optimizes the customer interaction. How do the organization's tools, practices, and processes add to these people's impact? How does the sum of the parts become larger than the individual parts?

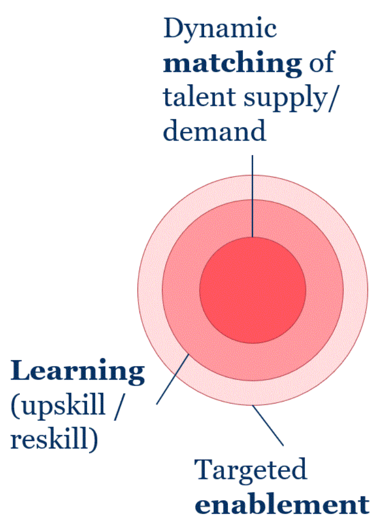
Let's peel the onion, a neural layer at the time.

The neural net *you* inadvertently designed

First, individual people owe their capabilities to their experience, and their impact depends on the **match** between those capabilities and the job at hand...

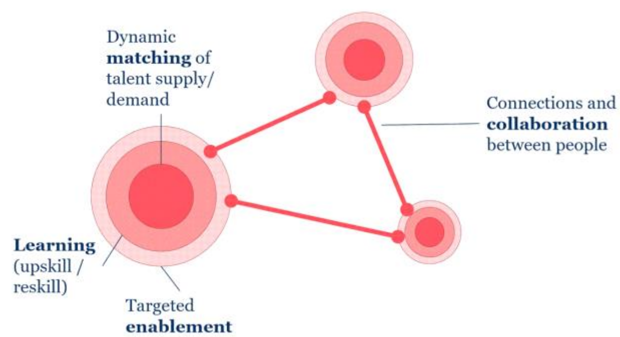


...then performance can be enhanced through **learning** and development (L&D) activities, which help people develop pattern recognition in the new environment.



...which is further **supported** in the flow of the specific processes they run, thanks to management support, documentation, etc.

Then, people **connect** with others to attack problems that they can't solve by themselves. Both L&D and enablement resources are provided by the organization, and those practices are among some of the best predictors of enterprise effectiveness[i]. But here's the important twist: those capabilities are *amplified by the connections* that people have, which help them complement their skills and use others as creative soundboards. The combination of individual knowledge (existing, new, and targeted enablement) with network connectivity generates collective intelligence that's superior to the mere sum of people's own intelligence and



their knowledge[ii].

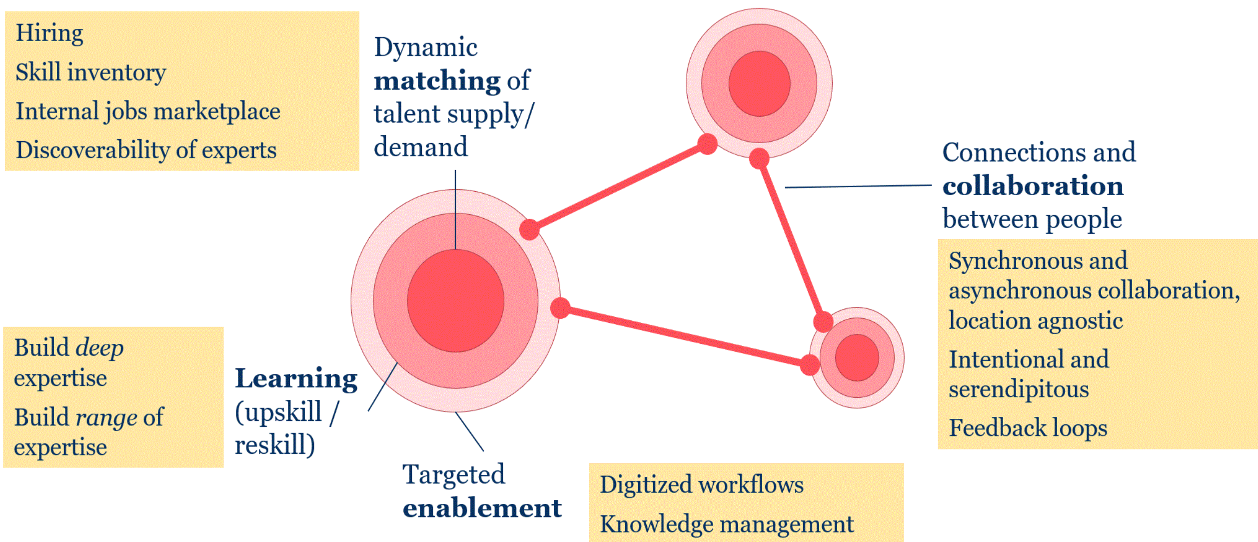
If this looks like a neural net not unlike our brain's, that's because it likely does some of the same, at a different level.

Designing and building your organization's collective intelligence

The result is the scope of work for CIOs and enterprise architects who need to design the "augmented collective intelligence" system with COOs, CHROs, and P&L owners. Look closely at the yellow boxes - that's technology and practices you've likely built organically. It is time to make them part of a cohesive plan.

The enterprise technology market is understandably involved in this, though often in a fragmented way, making enterprise architecture more challenging.

All of these solutions augment collective intelligence and should be part of enterprise-wide organizational design efforts. The prize is competitiveness in unpredictable times.



Some are **established capabilities with modern data components** (e.g., skill inventories or learning—e.g., what we built using collective intelligence or what Workday embeds into its ERP SaaS). Some are **emerging capabilities** that enterprise vendors such as Workday or Gloat are focused on—for instance, internal jobs marketplaces or AI-enabled chatbots that monitor employees' engagement. Some are **net-new things** that fall outside of traditional categories, like virtual watercoolers based on dynamically generated network analysis.

AI's Human Side

Published: October 26, 2019

In 1998, two years before the carnage of the dot-com bust, two Stanford Ph.D. students presented an interesting scientific paper. It contained a big idea that made their company one of the most valuable in the world. Their name was, obviously, Sergey Brin and Lawrence Page, and their company one of the then-many search engines – the one that finally gained one of the most lucrative competitive positions in history. The world has since been entranced with their superior digital prowess, and especially their mastery of artificial intelligence (AI) – together with the mound that would create to protect their market position and their valuation. Senior leaders and consultants have studied and attempted to emulate those strategies and practices – with, to date, comparatively scant results.

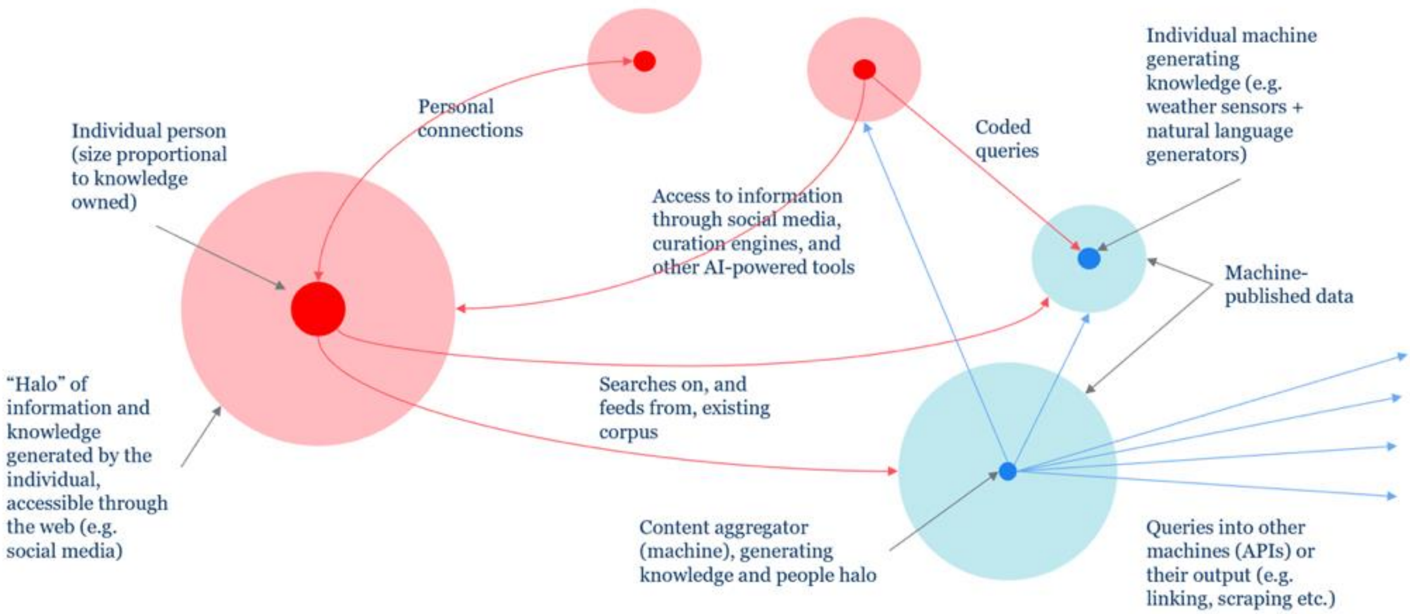
There's however another part of that story, lost in the mainstream AI hype, that could help "the rest of us" harness some of the same power. Let's go back to 1998.

The other part of Google's AI

Google didn't invent the search engine but did achieve two things that changed the world. First, its queries became more relevant through a new way of ranking the results pages. They did so by measuring those pages' "centrality" in a broad *network* of websites, the web pages created by thousands (and later millions) of people that reference each other through hyperlinks. Second, they were quick to embrace and transform the new generation of online advertising, which relied on those queries' relevance to target users more accurately and, as a result, more valuably. The combination of the two was one of the most powerful business model innovations of the last century.

What most of us don't realize however, is that Google's power comes from masterfully using ("organizing", in Google's parlance) the fruits of the intelligence of billions of people: the knowledge they create, and the choices they make when they browse. That is, Google's business model wouldn't exist without the intelligence of knowledge producers, curators, and users – harnessed in ways unthinkable just twenty years ago, and continuously growing (today, almost 2 billion websites exist). In so doing, Google also contributes to that intelligence, by enabling the world to retrieve knowledge in a sort of collective, comparatively frictionless "remembering".

The explosion of other social media, from blogging platforms like WordPress to Snapchat, has fueled the "creative fire" of millions of people. While much of that creation is of dubious intellectual value, interesting ideas often stem from these environments. These technologies are increasingly able to collectively “sense” the environment: From the Arab Spring to breaking news and crises, a massively decentralized network of sensors (the majority being simply people with smartphones) has emerged, with its flow of ever-fresh information.



Despite the hype, many internet giants’ strengths don’t just come from the intelligence of their AI, but also from their use of the cognitive power of billions of people who generate information and decide to listen to one another in very specific ways. Trillions of microevents, a sea of new information – all made by human choices - every day. (To us, individually, those choices aren’t worth much, but they make a huge difference to an advertising machine and are worth trillions of dollars of market value.)

These examples matter to most people. They can inspire us to more deliberately leverage the intelligence *networks* that surround us—within and outside of our organizations.

Look at the picture below. Neural-like networks like those, enabled by AI, now span vast numbers of sources of knowledge, especially people but also machines. They weave those nodes together and spread their ideas thanks to web-enabled hyper-connectivity, generation of sensor-based data (from weather to stock inventory to citizen’s warnings on Twitter), curation of content, display of that content to relevant parties through prediction (think social media choosing what you’ll like to see), and connections (web publishing, synchronous communications, etc.).

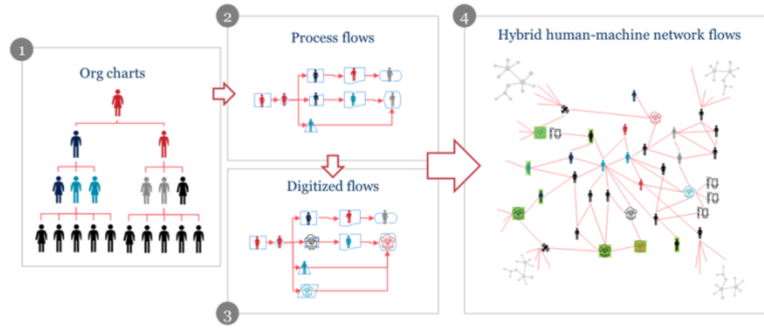
The third neural net

Within organizations, that's an explosion of knowledge available, absorbed and filtered by organizational networks now solidly wired through the likes of Outlook and Slack. That knowledge is then amplified and evolved by social networks and is simultaneously immersed in – meshed with - continuous streams of other ideas curated by AI-based algorithms. Not coincidentally, one of the fastest-growing spaces in the field of analytics is that of "knowledge graphs", whose biggest advantage is to document and process relationships similar to the ones displayed above.

That's here, today. That's why intelligent networks, made of large numbers of people *and* AI-powered machines, could be a new organizational design ready for widespread adoption. They can help many leaders, from CEOs to middle managers, from centers of excellence to movement organizers, harness the full collective cognitive power of their organizations – to generate and implement stronger ideas, and adapt more swiftly and effectively to fast-changing conditions.

Consider the below: what happens when we *add* to the traditional management methods (in figure 1, 2, and 3) the ability to *deliberately orchestrate* networks that span across organizational boundaries?

Thanks to AI, that can be done today. What could our organizations become by



fully using people-machine networks made more intelligent by **AI's "four C's"**, i.e. its ability to exponentially improve the following four things?

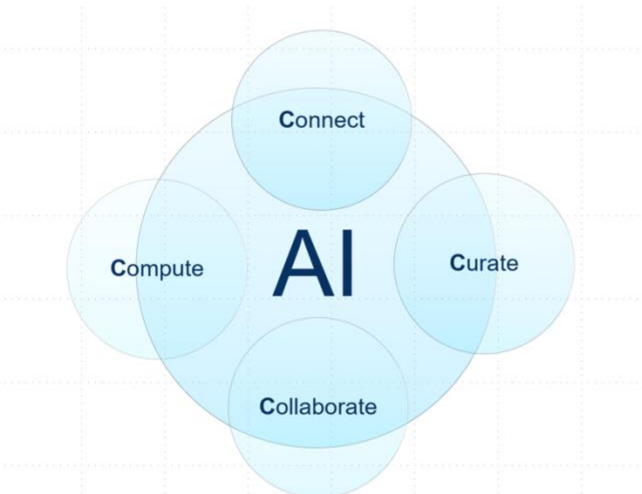
1. **Connect** entities (people, and machines – by, for example, helping pinpoint the right nodes in the network and making them discoverable to each other through search);

2. **Curate** knowledge (for instance, semantic searches and computer vision that identify the most relevant content, and cluster it for people to process it more easily);

3. **Collaborate** across those entities now enhanced by the new knowledge (for instance, natural language processing that automatically translates content or machine learning that optimizes video and voice transmission);

4. ...and **Compute** any other prediction (for instance, to determine which participant in the network is worth rewarding, or

other machine learning algorithms
to detect spam and fake inputs)



We don't know the answer for sure, but work from MIT and others over the years hints at the possibilities: the creation of a networked, connected intelligence that could make our organizations smarter and turn them into intelligent *systems* able to sense, create alternatives, act, and learn—that is, adapt—over time.

Stop Working Like It's 2019

Published: November 24, 2020. It is interesting to re-examine the thoughts that the early pandemic time ignited. Most of the organizational principles still apply – but we now have better technologies and practices to implement them.

After months away from offices, it is amazing to see the massive adoption of new tools, from Zoom to Slack to Teams and many others, with significant benefits that will outlast COVID-19. One would wonder why we didn't collectively use more of these tools before. Yes, much of it is exhausting, and some of it is clunky, which prompts some to say, "It is better to be in an office."

There's reason to believe they're missing a good part of the point. The reality is that many of us are still collaborating through digital tools the way we did in physical meetings. We're having more meetings, fast, but there are still lots of meetings and emails. Let's face it: **we are often force-fitting old practices into new tools.**

Virtual-native work

So here's what I suggest: redesign parts of your work, and especially how you collaborate, in a way that's "**virtual-native**". **Try using the tools for what it would have been *impossible* to do before**, not for what you *were* doing before. For example, try:

- **Asynchronous meetings.** Meetings should be a process, not an event. Not everything needs to happen on a punctuated event like a call. Try using Teams, Slack, or the like to get people to collaborate prior to *and* after the physical meeting. That makes things faster and doesn't force everyone to be on the same call at the same time.

- Use extensively **meetings recordings and related transcripts**. Doing that on Zoom or Teams for instance is a breeze (thanks to cloud computing). Once posted on some of the threads mentioned above, they become a knowledge library and a minutes repository. They can be scanned much faster than sitting in a full-length meeting.

- **Go "co-editable first."** Creating documents, whether slide decks or word documents, should increasingly be done on a central cloud-based repository. First, the structure should be set, and then contributors should build the document independently while giving each other extensive feedback. No more documents should sit on someone's hard drive and be sent around by email.

- **Ideation at scale or feedback.** Get large groups of people to generate many ideas in a short time, both synchronously and asynchronously, by using some of the native functionality of Teams or similar, or specialized tools such as Miro, Mural, Slido, etc. When done right, it feels like magic.

- **Global creative flash mobs.** This is a derivation of the above, with a twist: because distance doesn't matter, whoever is awake at that moment in time can contribute. That opens up the incredible possibility to get the right people (and more cognitively diverse people) within a matter of hours.
- "**One whiteboard a day** keeps obsolescence at bay". Seriously, the number of people using a whiteboard since evacuating offices has plummeted. That's a really bad thing. Whiteboards are a phenomenal problem-solving tool and a great storytelling method. Whoever owns a tablet (e.g., an iPad a touch street PC) and an inexpensive rubber-pointed pen has no excuse.
- **Internal crowdsourced predictive markets.** Interest in these things has ebbed and flowed, largely because some were misused, and managers didn't really know how to deal with the process change management required. Things like Unanimous.ai and many others are worth a look again.
- **Reinforce the social weak ties** that bind people and strengthen culture. Go out of your way to reconnect, and have your teams reconnect, with people they used to bump into, but they don't anymore because canteens and water coolers are gone.

Harness the network's intelligence

But if you're brave there's a lot more, which some daring leaders, all the way to the CEO, should really try out with their organization: design the workplace, digital and otherwise, to harness your **network** of smart people through intelligent machines...not the individuals in isolation. Four key tenets below.

a. **illuminate the network**: intentionally map the "nodes" in the organization - especially influencers who tend to be invisible in traditional org charts

b. **incentivize the nodes**: give them incentives to give back to their entire networks - for instance, to help curate relevant and fresh knowledge with the help of some central resources

c. **power the "knowledge intake"**: ensure that diverse sources of external and internal information can be mined systematically

d. **perfect the collaboration platform**: including some of the practices described above. But that's just a start.

All of this doesn't mean that offices won't exist in the future. But for many, they will be an *emotion*-focused place to recharge social capital, team energy, trust, and engagement. They may not be the primary place to do heads-down work, and they won't become a pretext to insulate people from their colleagues and partners who work elsewhere.

Lessons from the past: how augmentation is hindered

Many leaders are somewhat suspicious about the current AI hype cycle as they remember the previous one. There is a view that the last wave of enterprise AI, started in the mid-2010s, failed to deliver on its expectations fully. Is that true? If so, or at least partially so, what are the technological, process design, people and organization, and other reasons for that outcome? And even more importantly, *what can we learn from it* to help us make better decisions in today's AI wave?

The last question is critical. Yes, there are differences between Predictive/Analytical AI (the majority of AI implemented between 2015 and 2022) and Generative AI: the former was more centralized, as it required specialized competencies to design, build, and run. The latter lends itself to more decentralized experimentation, as workers can touch it and use it easily. However, there are also significant similarities, especially as one tries to scale Generative AI's use, with its attendant infrastructure, cost, rigor, and operational excellence (including reduction of error rates and process design) - in other words, work across the entire operating stack. It is worth comparing the two.

Did the “Classic Enterprise AI” Wave from the Mid-2010s Fail to Deliver Fully?

Yes, *on average*, in broad terms, and within the time horizon that was initially expected. Indeed, many, perhaps most, enterprises struggled to realize the transformative potential they initially envisioned. Despite heavy investments, many companies could not scale many use cases beyond proofs of concept, integrate AI into core operations, or achieve enduring strategic impact. The gap between early hype and real-world outcomes underscored significant challenges that became apparent only after the wave’s initial enthusiasm subsided. So, is history repeating itself?

That would be rushing to the wrong conclusions. Averages hide much variance, and variance is where the learnings (and earnings) are. Better said: the AI age of the 2010s led to a dramatic emergence of a few companies that outcompeted the rest, permanently altered their markets, and created disproportionate value.

The AI age of the 2010s led to a dramatic emergence of a few companies that outcompeted the rest, permanently altered their markets, and created disproportionate value.

The increasing dominance of these tech giants is evident in their growing share of the total U.S. market capitalization: In 2012, they accounted for about 6% of the U.S. market cap, and by November 2024, their share had risen to 27%. The collective market capitalization of the

Magnificent Seven (Apple, Microsoft, Alphabet, Amazon, Meta, Tesla, and now including Nvidia) has grown dramatically: In 2012, their combined value was approximately $1.1 trillion; by November 2024, this value had surged to $17 trillion. Services companies that caught at least part of that action (say, Accenture) also did well.

Additionally, a minority of organizations—particularly those in technology-intensive industries, fintech, and innovative high-tech firms—achieved tremendous value and competitive gains. Not coincidentally, most of the value created in the stock market in the last fifteen years has accrued to such companies.

That is not just because their "vibe" feels more contemporary but because their growth is more substantial, and they seem to skate towards "where the puck will be" - and the corresponding value pools. Even accounting for some valuation's "irrational exuberance," a clear value shift is underway toward companies that have embraced AI at their core.

Apart from them, certain tech-forward and innovation-driven, yet not AI-native, companies sidestepped these pitfalls by deliberately addressing the known challenges.

Certain tech-forward and innovation-driven, yet not AI-native, companies sidestepped these pitfalls

Some financial services firms, including Fintech and investment banking, and selected others (like online banking) reaped results. The IT service landscape has changed dramatically. Some professional services and management consulting firms (BCG, McKinsey) harnessed the new wave profitably through service portfolios and how they deliver that work. Among others, especially those whose operations are heavily physical, Walmart stands out as a company whose leadership position could have been heavily eroded by the rise of e-commerce and the power of Amazon and others and instead managed to innovate enough of its core to ride that wave and maintain a solid position (although its value as a company is a third of Amazon's). Many other "traditional" companies have outcompeted their rivals by improving their operational performance through AI, whether through better customer support, supply chain, or sales & marketing.

At the time, we were told that companies that didn't invest quickly to undergo the learning curve of deploying AI would never be able to catch up because AI is a "winner-take-all" game with steep experience curves. That might not have been true for all industries and segments, but it was very much true in those where some companies did meaningfully embed AI in their core.

To be clear, quite a few people called out the potential pitfalls (I was **one** of them).

It is worth revisiting them because they hold many keys to tomorrow's success.

What explains the failures? Six pitfalls of the “classic enterprise AI” cycle that the non-AI-native “rest of us” should stay away from

They’re not just about the algorithm and the data; they’re not just the CIO/CTO’s responsibility. In many respects, they are a collective failure across the operating stack: people, policy, processes, applications, algorithms, tech infrastructure, and data. They’re more strategy, organizational design, and leadership than is typically assumed.

Let’s dive into each to understand what went wrong and how focusing on them can help with today’s challenge.

Technological and Data Maturity

Immature Tooling and Ecosystems: The foundational machine learning frameworks advanced rapidly, but the supporting ecosystem—robust platforms, monitoring tools, and deployment pipelines—lagged. Without mature MLOps practices (automated CI/CD/CT for models), organizations struggled to maintain model accuracy, continuously improve solutions, and ensure proper governance and compliance.

Data Quality and Accessibility Issues: AI requires large volumes of clean, consistent, and accessible data. Enterprises often faced fragmented, siloed data trapped in legacy databases and outdated architectures. Data engineering efforts consumed

Area	Root causes 2015-2022
Technological and Data Maturity	Immature Tooling and Ecosystems Data Quality and Accessibility Issues
Market and Vendor Ecosystem Factors	Overhyped Vendor Claims Cost Overruns and Uncertain ROI Timelines
Legacy Systems, Processes, and Human-Centered Design Failures	Entrenched Legacy Technologies and Practices Human-Centered Design Practices Not Fully Adopted
Inappropriate Innovation and Development Approaches	Conventional “Design-to-Build” Mindsets Insufficiently Adaptive Innovation Frameworks Proofs of Concept (POC) Stagnation
Organizational, Cultural, and Leadership Factors	Skills Shortages and Siloed Teams Cultural Resistance and Underinvestment in Change Management Insufficient Executive Championing and Strategic Alignment
Governance, Compliance, and Ethical Complexities	Insufficiently Robust Governance and Model Lifecycle Management Regulatory and Ethical Uncertainties

disproportionate time and resources, hindering model development and deployment progress.

Market and Vendor Ecosystem Factors

Overhyped Vendor Claims: The vendor ecosystem frequently promised quick wins and plug-and-play AI solutions that downplayed the complexity of data preparation, integration, and organizational change. Enterprises that trusted these hyperbolic claims encountered disappointment when reality failed to match marketing rhetoric.

Cost Overruns and Uncertain ROI Timelines: Many early AI investments were significant without delivering near-term payoffs. Management's patience often ran thin as budgets ballooned and ROI remained elusive, prompting some organizations to scale back their ambitions prematurely.

Legacy Systems, Processes, and Human-Centered Design Failures

Entrenched Legacy Technologies and Practices: Instead of reimagining their IT landscapes, including new AI and non-AI technologies, companies frequently attempted to layer AI onto legacy systems and processes never designed for agility or scalability. Without modernizing these foundational elements, AI solutions were left to contend with brittle integrations and technical debt.

Human-Centered Design (HCD) Practices Not Fully Adopted: Although many spoke of putting the user at the center, and despite the design-thinking hype, few organizations truly embraced design-for-transformation principles. HCD was often treated as a check-the-box exercise rather than a genuine effort to co-reimagine, not just chisel workflows, interfaces, and experiences aligned with AI's capabilities. We focused these methods mostly on known-known (known set of issues, known set of solutions) problems instead of wielding them to explore the unknown-unknown space. Consequently, AI was forced into existing operational molds, limiting its transformative impact.

Inappropriate Innovation and Development Approaches

Conventional "Design-to-Build" Mindsets: CIO- and CTO-led teams often applied linear innovation processes best suited for stable, well-understood technologies. The introduction of Agile made the software deployment less linear, but the issue was upstream from that. As a rapidly evolving frontier technology, AI demanded iterative, exploratory methods engaging the whole enterprise that adapt as capabilities and market conditions shift.

Insufficiently Adaptive Innovation Frameworks: With the technology frontier constantly moving, relying on standard initiative-prioritization practices proved ineffective. More suitable methods would have segmented investments across horizons (H1 for near-term gains, H2 for emerging capabilities, H3 for long-term breakthroughs), used clear stage gates, and encouraged experimentation, pivoting, and continuous learning. This is what innovation teams do, but they should have been more deeply integrated into the CIO/CTO and Transformation process and its resource allocation.

Proofs of Concept (POC) Endless Purgatory: Many AI projects never evolved beyond the pilot phase. Without a solid innovation - not implementation - methodology for moving from idea to POC to full-scale production, initiatives remained as lab demos or small-scale experiments that failed to deliver enterprise-level returns. Absent plug-and-play, high-accuracy, cost-effective, and scalable AI solutions, the cost of running the operating stack (not just the IT one) remained prohibitive, and many efforts realized that too late. Unable to pull the plug and write off the cost, many stayed stuck in purgatory, which also crowded out other, possibly better efforts.

Organizational, Cultural, and Leadership Factors

Skills Shortages and Siloed Teams: Data scientists, ML engineers, and product managers with AI expertise were scarce. Even when hired, they often operated in isolation from business units or end-users. Such silos produced solutions disconnected from real needs and hindered the cross-functional collaboration necessary for meaningful AI integration.

Cultural Resistance and Underinvestment in Change Management: Implementing AI frequently required rethinking job roles, decision-making processes, and long-standing workflows. Many enterprises underinvested in training, communication, and culture-building initiatives. This lack of support made employees skeptical, resistant, or unprepared to work alongside AI-driven systems.

Insufficient Executive Championing and Strategic Alignment: Weak or inconsistent executive sponsorship made AI a side effort rather than a core strategic initiative. Without top-level support, dedicated funding, a clear long-term vision, and integration into corporate strategy, AI projects struggled to secure organizational buy-in and momentum.

Governance, Compliance, and Ethical Complexities

Insufficiently Robust Governance and Model Lifecycle Management: MLOps and model governance frameworks were underdeveloped, leaving organizations unable to track performance consistently, maintain compliance, ensure fairness, or address ethical concerns. Without strong governance, AI initiatives lacked the rigor required for stable, long-term operations.

Regulatory and Ethical Uncertainties: In regulated industries such as healthcare or finance, unresolved questions about data privacy, fairness, transparency, and accountability created hesitancy. Unclear or evolving regulatory landscapes often forced lengthy deliberations and revisions, delaying deployments and slowing innovation.

Interestingly, while we have improved at all of these, the ante was upped on us - with more complex, faster-evolving, and powerful technologies - while some of the bad taste in (aging) executives' mouths increased.

While we have improved at all of these, the ante was upped on us

The result? Many still sit on the fence, waiting for more proof, stability, and resources.

Learning from the past in the AI age is not an oxymoron

The "classic enterprise AI" wave from the mid-2010s did not fully deliver on its lofty expectations because many organizations were overwhelmed by the complexity and breadth of changes required. Lacking appropriate innovation frameworks, struggling with immature technology ecosystems, being hampered by legacy processes, and falling short in skills, governance, and cultural readiness, many enterprises saw comparatively limited returns—outside of those areas, like videoconferencing, where AI simply became "ambient." However, a minority that methodically addressed these challenges achieved huge, world-changing benefits.

Surveys and analysts tell us that most AI experiments fail, and many critics say, "I told you so." However, a 70% failure rate before or at proof-of-concept is to be expected for this type of technology. The first wave of enterprise AI was the same. Remember that *a 70% failure rate doesn't mean we would be wasting 70% of the money*.

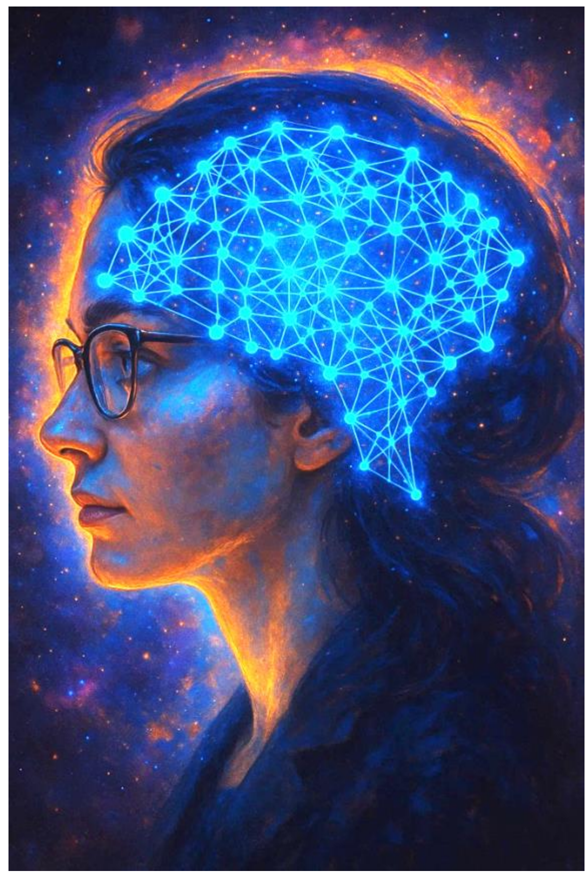
A 70% failure rate doesn't mean we would be wasting 70% of the money

Anyone who does innovation for a living knows that the early part of the innovation funnel is an attempt to "fail fast and cheap." But treating innovation-grade technology like you would treat a regular tech investment is doomed to fail - and unnecessarily hurt.

While history doesn't necessarily repeat itself, it sometimes rimes. In this case, the underlying principles of technology and organizational and process design still apply to the new conditions. These lessons must now inform the approach to the latest wave of AI, including Generative AI and its potentially very distributed, system-wide impact.

We know better today, and that should give us confidence to step up our innovation efforts.

Improving Innovation and Problem Solving – with Augmented *Collective* Intelligence



Problem-Finding AI Agents and Exponential Serendipity

2025 is the year in which we remembered AI agents moving into production in many places, mainly to tackle initially small but increasingly meaningful *problem-solving* tasks. However, an essential part of their work happens *upstream* from problem-solving. That is, *problem-finding*. Problem-finding involves identifying the *why* and the *what* before working on the *how*. In many cases, this is at least as important as solution-finding, and quite often, it's what senior leaders must focus on the most.

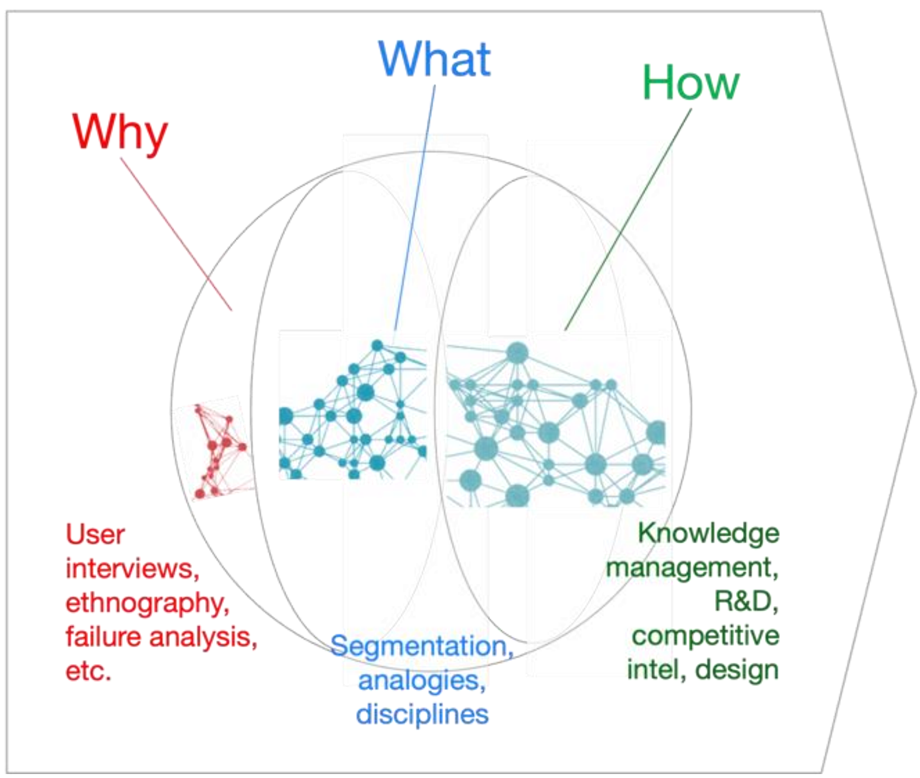
Interestingly, it is also where most people feel that machines, including AI, have the least capabilities - a feeling that Pablo Picasso encapsulated well when he said, "Computers are useless; they can only give answers." But - should that be true today?

Pablo Picasso said, "Computers are useless; they can only give answers." Is that true today?

Beyond the concept and intuition, I want to propose an architecture for doing this at scale—one that can be industrialized and yield exponential results using technology we already have and, indeed, the technology we'll have in 2025.

Ideas—including those related to problems—have a structure that machines can interact with

The core notion here is that ideas have structure (morphology). They can be divided into different parts and connected. Humans sense those components intuitively in their minds. Technology can do that via knowledge graphs or other representations that AI could generate if enabled and asked to. Then, when exploring an idea—its why, the what, and the how—we can ask AI to analyze it deliberately, including the structure of ideas that relate to *problems*, not solutions.



Even if AI machines handle some of it in their answer (at inference time), this additional reasoning step and the tokens involved can yield tremendous value in finding more accurate and conceptually non-trivial solutions.

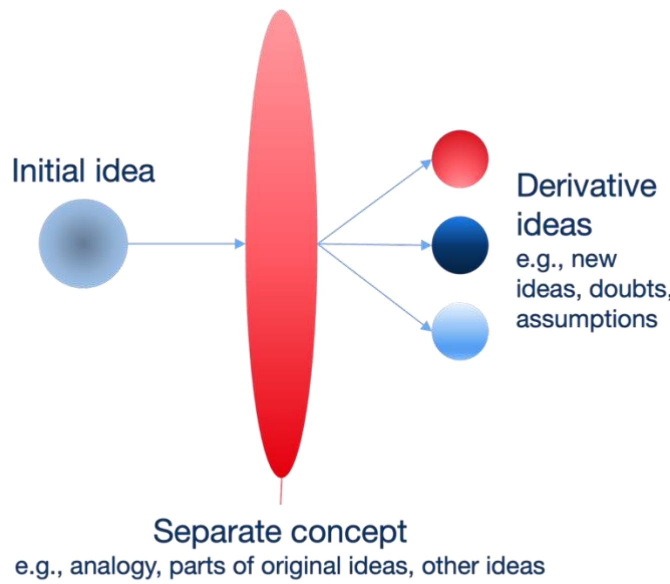
Today, these exercises are carried out in workshops where participants give each other ideas or systematically seek them through some form of "information and knowledge feeder," a curation engine (the like of Perplexity, of Gemini DeepResearch.) But all of this is slow, manual, and suffers from the limitations of human's "field of vision" which makes the serendipitous discovery of new ideas clunky and inconsistent.

What can be done today? Likely, if what we seek is novelty, "extrapolation", we can't just ask AI point blank. So let's take a step back and reflect on the appropriate thought process through something that oddly resembles chemistry and physics.

Forcing idea collisions

In a previous essay called ***AI Can Ideate Harder***, we saw that we can use AI-enabled ideation processes to force ideas and their components to collide with others deliberately. In that process, we also use lenses—conceptual frameworks based on solid human reasoning encapsulated in a framework—and smash ideas against them. We do that both to find solutions *and* problems by exploring and exposing different dimensions of the problem to discover new angles of attack.

We can use AI-enabled ideation processes to force ideas and their components to collide with others deliberately. We do that both to find solutions *and* problems.



A great way to do this traditionally is to seek out interesting people and *their* ideas and then compare ours with theirs. By the end of that conversation, we better understand the problem at hand (the Why) and the potential categories of possible solutions (the What). This then informs the downstream problem-solving work.

We do this routinely in our offices, in meetings, at water coolers (physical and possibly virtual), at conferences, on social media, and so on. In a way, our societies thrive because there is this "perpetual motion machine" of idea collisions across our networks, and those ideas get harvested by organizations—companies, academia, and even entire markets.

Enter Problem-Seeking AI Agents

What's exciting is how AI can exponentially amplify these processes in non-trivial ways. Given enough resources and guidelines (including ethical, bias, and intellectual-property), AI can relentlessly canvas people's "halo"—the corpus of knowledge surrounding them. Machines can identify interesting individuals, examine their ideas (in this case, their challenges), and "smash" ours against theirs. That works very well for problem-finding too. As a result, AI can serve us knowledge that is a net-new, **relevant addition** to what we already know instead of forcing us to wade through many things before we find something really accretive.

Machines can identify interesting individuals, examine their challenges, and "smash" our capabilities against theirs to find meaningful recombination. That works very well for problem-finding too.

In problem-finding, an AI agent could locate people whose ideas benefit from our broader capabilities and do a first round of iteration to discover intersections, either existing or potential (that is, requiring additional work), before handing them over to humans and their other AI tools (such as **scaffoldings and exoskeletons.**)

A first simple application that could make each of us a bit of a superhuman problem-finder: an AI agent sifts through all relevant newsletters, identifying the key trends that others are trying to solve and giving you not just a deduplicated digest of the zeitgeist but a "net-new" one that has filtered out anything that you know already, hence increasing the output **usefulness** and reducing your cognitive effort.

Now open the aperture: imagine you run an innovation ecosystem and know (your AI's corpus knows) the talent and technologies it typically comprises. You can now send AI agents into the halos, the knowledge spaces of potential target companies and those who lead them. The AI can find combinations between what those companies seek, their unresolved needs, and what my ecosystem can potentially offer with additional innovation efforts because the "why" and the "what" are mapped thoroughly.

You can extend the examples to many other spaces:

In **business-to-consumer**, your problem-finding AI can continuously canvas the synthetic personas of target customers, especially as they dynamically evolve because of market trends. In the media industry, an AI could scan online streaming habits, fan forums, and influencer trends to uncover unserved audience interests. It might detect, for instance, a rising fascination with eco-thriller narratives, suggesting a new genre mashup that existing studios or content creators haven't explored yet. By reviewing consumer transaction data, credit bureau updates, and macroeconomic indicators, an AI could surface potential underserved financial segments—like gig economy workers lacking stable cash flow options. By analyzing electronic health records, research publications, and local environmental factors (e.g., pollution levels and dietary trends), an AI agent can detect emerging disease clusters or unaddressed care gaps.

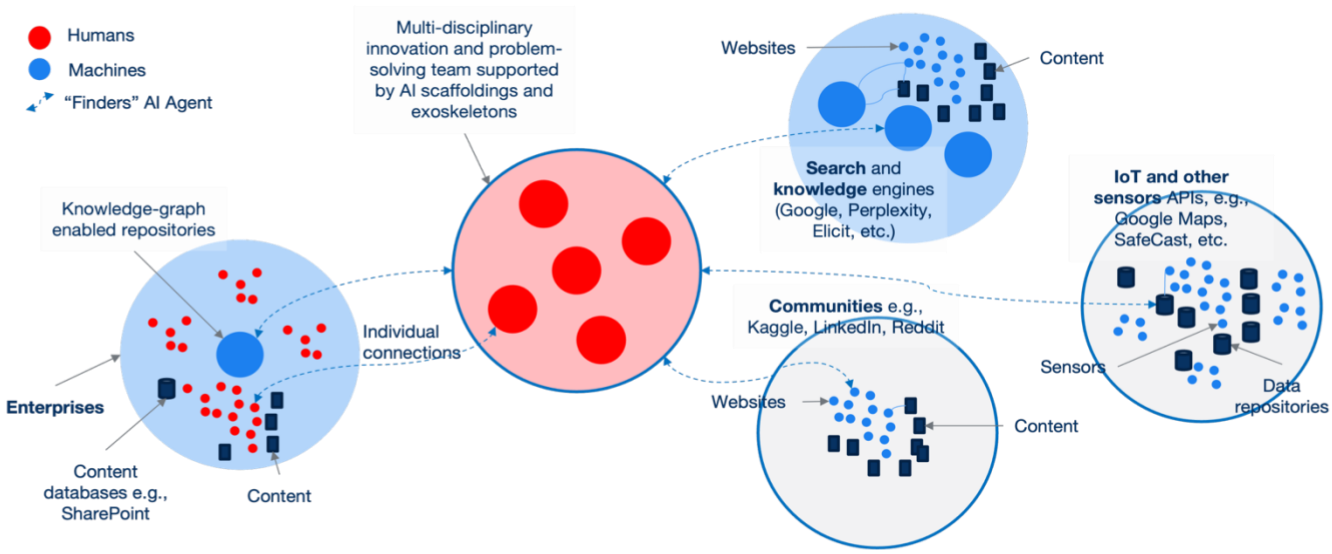
In **business-to-business**, your problem-finding AI interacts with the body of knowledge accumulated around target clients, both the organizations (for instance, through their product/service portfolios and their earnings calls) and the buyers themselves (looking at their public statements)

Your problem-finding AI can also scour machine-generated data, such as those from IoT sensors (think weather data), and combine it with others. For instance, insurers detect weather and home-improvement trends to address climate-related risks in the residential market. Problem-finding AI can also sift through global shipping data, real-time traffic feeds, and weather forecasts to identify new types of bottlenecks well before they happen.

New **scientific and patent data** can also be engaged with, whose "why" reveals the emergence of new partial solutions that hint at new problems. For instance, decades ago, Corning's Gorilla Glass could have hinted at new user interface designs. Or, today's lightweight edge AI hints at new, solvable problems, from industrial logistics to environmental monitoring. An AI agent can survey scientific studies, sustainability metrics, and community impact reports and discover that specific recycling initiatives are failing due to poorly chosen plastic types.

And, of course, the sky is the limit. Imagine how this can support strategy (also including M&A) and finance teams, as well as their CEOs. But they could also enable public bodies to identify upcoming challenges and inform their scenarios.

The chart below illustrates a possible architecture of these systems.



The result is a better understanding of which problems are worth solving—problems that are both desirable and whose solution is increasingly feasible.

Agent, meet my agent—and talk through my stuff

To recap, a problem-finding agent is an AI-driven system that [1] scours relevant data sets (knowledge halos), [2] identifies new or unmet needs, [3] cross-references areas that are attackable by the categories of solutions that exist (even if the exact solutions don't), and [4] presents them to humans for deeper evaluation.

Conventionally, this kind of synergy happens through people who meet each other—either systematically or serendipitously. But now, machines can go and "meet" with the knowledge base of those people. In a not-so-distant future, these machines could even meet with other people's AI agents, allowing for an initial and thorough scan of the potential for intersection and identifying additional problems to be solved. Once they find enough overlap—enough interesting problems that should not only be solved but can be solved—then humans can meet, possibly assisted by machines with a thorough context.

There has been much discussion of agents independently engaging other agents. This is one of the more straightforward use cases I can imagine, as it benefits from scale and is low-risk.

This is how companies can start:

1. in concert with your technology team, business teams should identify the AI agent capabilities you can deploy in the next 3 months. Don't undertake anything that can't be done in a short sprint

2. design AI-finding agents human-centered, by involving users early in determining what would be truly desirable to them (use cases, user experience, human-in-the-loop feedback) and what

data (knowledge bases across silos) you can access, and avoiding unnecessary over-specs

3. start small with a proof of concept, but don't try anything for which you don't have a path to scalability

Individual and organizational resilience hinges on detecting inflections early in an increasingly fast-moving world. AI problem-finding agents can help there. In the process, we might heed Picasso's concern and use computers to do more than just give answers.

If all of this sounds like the inception of an Iain M. Banks novel, it may very well be.

And the pieces are in place for it to be a reality.

GenAI Must Ask Questions, Not Just Give Answers

Some of Generative AI's limitations, especially the models that are currently widely available, stem from a very simple thing: they're calibrated for the "instant gratification" of their users. That's a real problem when you're trying to solve *complex* challenges that have defined "right/wrong" answers, and that's one of the reasons why sometimes language models provide well-structured, polished yet ultimately "middle-of-the-road" and unimaginative answers or even grossly inaccurate ones. There are ways to address this challenge, and they have to do with a rethink of what we mean by "generative".

The value of thinking slow

GenAI tools often remind us of over-eager interns who want to show off, so they are perceived as smart. (To be sure, many other professionals fall into that trap: trying to impress colleagues by giving fast answers.) Our own, human, fast-answer mechanism, which Daniel Kahneman dubbed "system 1" thinking, is often misguided. We all - humans and machines - need "slow" thinking when looking for truly interesting answers. And that, I argue, starts with AI asking more questions to humans.

Three examples illustrate the potential of this intuition.

First, research conducted in mid-2023 by Harvard and BCG showed that large language models (LLMs) could already improve the output of junior consultants on tasks that don't require much proprietary and private context, such as new product ideas for a business-to-consumer market segment. However, they struggle more with providing cogent company strategy, whose quality heavily depends on the understanding of the broader company context.

A second set of examples comes from healthcare. While LLMs are already showing remarkable ability to provide answers when given thorough anamnesis, the reality is that most patients are unable to give a good enough set of symptoms. A large part of the role of doctors is (or should be) to also ask questions, and get more input for the diagnosis. An LLM could be instructed to do so, and even more so when using knowledge-graph databases to look into adjacent fields - which is something that even doctors struggle to do. Google AMIE does some of that, with solid results. Perplexity.ai and its CoPilot mode ask specific questions about the cause of health ailments before providing healthcare solutions.

A third example comes from education. Khanmigo, the tool built by Sal Khan (of Khan Academy) in collaboration with OpenAI, uses a sophisticated Socratic flow, to ensure that children develop a full understanding of the knowledge they absorb by being asked to complete and build onto the progressively more complete inputs that the machines provide, instead of just passively memorizing answers.

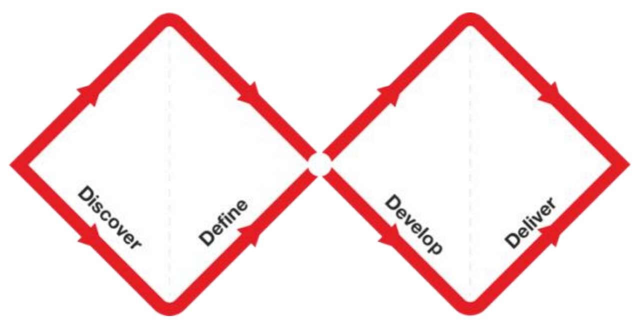
To partially summarize these findings, let's focus on context, computational efficiency, and semantic vs symbolic reasoning.

- AI machines, and in particular GenAI, often lack **context**. While their impressive confidence may tempt us to think otherwise, they can't read our minds, and many of us are tricked into thinking that they know more about our specific context than they really do.
- They also try to find answers in a **computationally efficient** way, which means that they won't necessarily comb through all the components of a problem, and won't recombine possible parts of solutions if they have a shorter and cheaper yet likely path to the answer.
- Finally, while they do seem to have some representation of the world that transcends pure language structures, they reason mostly **semantically**, which means that they currently struggle more than humans with symbolic reasoning and respective abstractions and generalizations.

And yet, abstraction, context, and taking the less-traveled path are exactly what problem-solving often requires when facing complex problems.

Three vectors for questions

Understanding the problem well is a big part of any solution. Asking probing questions can increase the aperture, and enable better focus - it is the core of the first diamond of the design thinking's double diamond, whose value is to identify the best angle for attacking a problem. But questions are extraordinarily useful in every other part of the ideation process too.



Source: UK Design Council, Wikipedia

No question, whether human or digital, is a bad question, as long as it leads to at least one of three things.

First, in these ideation processes, the **interplay between diverse participants** makes the difference between success and failure. Therein lies the opportunity: augmenting our collective intelligence by **amplifying the diversity of views** -

through better questioning facilitated by smart machines. AI can ask many types of questions, using one of the many frameworks built by humans that embed logical structures - for instance, Socratic and other logical thinking methods.

Second, additional possibilities stem from using machines to **radically open up the design space**, by asking humans (or other machines) to look at the problem through the lens of distant analogies (e.g. “What communities, such as Wikipedia, look like good analogies to healthcare knowledge management?”); or using ideas from other spaces, such as what Markus Buehler at MIT recently did when using GenAI for investigating engineering materials properties (supersonic fractures) through the lens of biology. Even when machines may not have intelligent answers, they can tee up creative questions that humans - and other machines - can then try and address.

Third, an emerging exciting avenue is through multi-agent generative AI: **multiple agents that ask questions to each other, and humans**. That process can be curated by humans who decide what types of agents are useful, or what threads of discussion are promising. Think for instance of a transformative idea for a large company, where much of the difficulty is in the change management and the acceptance by employees. What about a tool that uses multiple personas to critique the idea, acting as a synthetic, multithreaded town hall? For instance, one can gather the (simplified, possibly stereotyped, yet readily available) views of junior and senior employees, of people based in different regions, and of professionals in different departments. Humans could decide where to enter that town hall, decide which personas to amplify, or complement those voices, and draw conclusions and iterate on the transformative idea itself - as well as preparing a precise and thorough change management plan, including for instance detailed stakeholder engagement strategy.

These insights led to research and prototyping of alternative human-machine interactions for innovation. In our work at MIT's Center for Collective Intelligence (MIT Supermind Ideator), and subsequently in others, we built machines that ask humans to refine the questions before ideation. Solver also adds frameworks from insightful researchers, and theories - that put a lens on a problem, and through that prism it allows us to ask more and more pertinent questions. These are just small, early prototypes hinting at the potential of a very large design space that will be no doubt explored further in the near term.

As a not-so-small aside these considerations should also remind us that is also dangerous to put humans into a position of dependency on the AI machine, as this might lead to atrophying core cognitive traits - such as symbolic and critical, logical thinking - that people have. Designing for active interaction between humans and machines is crucial to maintaining the vitality of human intelligence.

Building questions-asking machines today

As we stand, we already have the practical means to build interactions with machines that make human-machine collective intelligence more powerful. While there is much more that could be done, a few simple ideas are below:

If you're using this approach for yourself or your team:

- Build a habit of never taking AI-generated answers at face value, and encourage the AI to ask *you* questions to help refine your thought process. For instance, systematically prompt the machine to ask you things like "What is missing?", "What else should I be thinking about in framing this problem, or solution?" or "How might this fail". AI can complement your views, or vice versa, but you always want to be on the receiving end of such questions.
- Make yourself and your team acquainted with the best-known logical thinking frameworks (you can even ask ChatGPT for them), and embed those methods systematically into the flow of your interaction with GenerativeAI. For instance, how would Christensen's Disruption Theory be applied to your problem? How would some Lean management principles (say, "The 5 Whys") shed light on it?
- Routinely submit your own work (e.g., drafts of presentations, solution ideas) to GenAI, asking AI to critique it by first asking questions to understand better what you're trying to accomplish so that the model can compare that with your artifact's content.

If you're building enterprise or client-facing applications:

- Deliberately build workflows that include steps where the machine asks questions to gather human input - whether upfront or as a refinement of the ideas generated by the AI.
- Consider adding multi-personas questions and critique (and even dialogue between the personas) into your workflows. This doesn't need to be a fully-fledged "Mixture of Experts Models" dialectic, as some of it can be readily obtained by engineering prompts, possibly even as part of OpenAI's GPTs.

Picasso said that computers are useless because they can only give answers. His point was also subtler: for true creativity and nontrivial problem-solving to occur, one needs to ask unusual, uncomfortable questions. Computers couldn't do that then. And mostly, we don't allow them to do that today either.

But with generative AI, they could. Generation is fueled by the right input and dialectic - it requires questions from us to the machine, but also from the machine to us. We just need to design the right interaction and workflows, and embed them into the flow of our work.

GenAI As Personal Problem Solver: A Case Study

In a day not too far into the future, your teams and you will routinely use AI tools as problem-solving assistants.

AI tools will help you identify solutions to "known knowns", problems whose perimeter and solutions are well understood by someone, somewhere. They will do so by helping retrieve existing, but not easy-to-find knowledge (think: how to get workers to optimize productivity in hybrid work).

They will help make progress on "known unknowns", the thorny problems whose boundaries are understood but whose solutions often need better collaboration across fields (think: process improvements for risk management processes in the presence of faster and more unpredictable risks, such as those stemming from climate change).

And they will likely put a dent into solving "unknown unknowns", the hairballs where the current solutions, and even the definition of the solutions, is currently suboptimal (think: establishing trust in AI-assisted social networks" or "mitigate the AI-induced labor unrest).

Here's a glimpse of that future, that is feasible today.

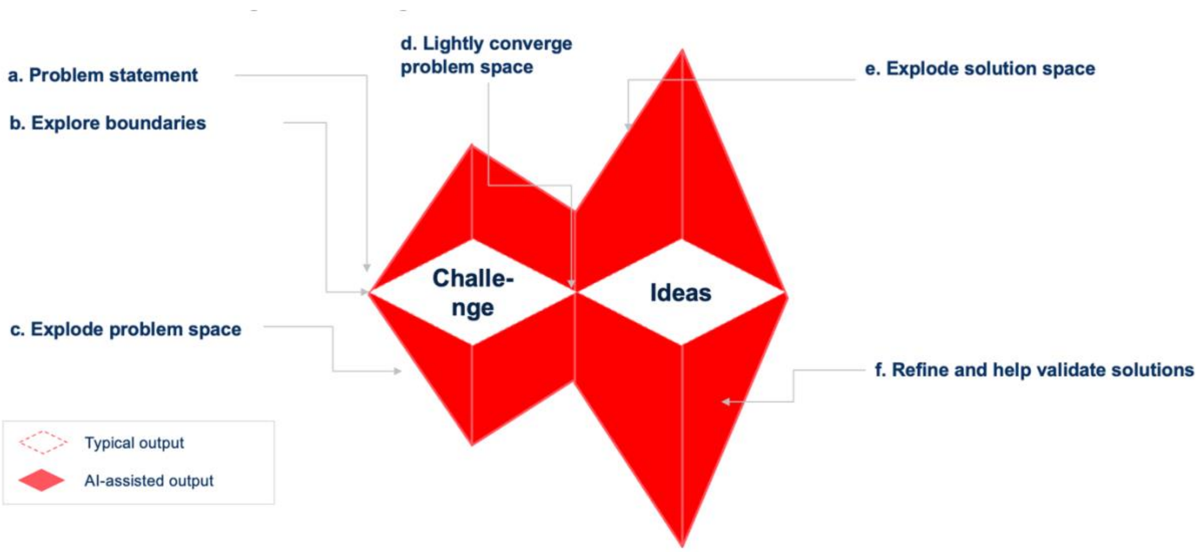
AI-augmenting innovation and problem-solving processes

My teams, colleagues, and I, at process and technology firm Genpact (NYSE: "G") and at MIT's Center for Collective Intelligence (see for instance the now-public MIT Ideator) have been working on making generative AI an effective enabler of problem-solving and innovation for since 2020.

To illustrate the potential of this new type of tool, I took two use cases similar to what many of us struggle with these days: preparing workforces for the impact of generative AI, and finding appropriate use cases for AI that are likely to lead to enterprise AI adoption.

Below are examples of simple, straightforward solution-identification workflows that took 10 minutes each to complete. (You find the outputs at the bottom. The prompts and code for chaining the workflow are not displayed here, as they're part of the tool.).

The tool is a Problem & Solution "Explorer", a type of AI-assisted "idea collider". Its high-level flow is simple and follows a typical problem-exploration / solution-exploration path, illustrated below.



Even when implemented simply, the value of properly configured AI-assisted workflows is that they amplify what humans can do in isolation, hence opening a larger space for exploration, conveniently, inexpensively, and scalably.

AI-assisted ideation process: two examples

Imagine that you have one of the two following challenges that you want to find solutions to:

A. **How to use AI in your business process.** Perhaps you're not convinced that you're using AI enough to improve your procurement function to its full extent. Sounds mundane? Procurement improvements go directly to an organization's bottom line, for instance; and procurement is heavily scrutinized for environmental and social sustainability purposes. A good example of a "known-unknowns" problem.

B. **How to prepare your organization, and its people, for the dislocation that Generative AI will create.** This is a more strategic task, typically approached with external consultants and involving senior leaders in the company. It is also a multi-faceted, unwieldy problem. In many respects, an "unknown-unknown" challenge.

You could read about the subject, involve some in your team, or get some experts in a meeting - but before doing that, and even while doing those things, you want to know enough to be able to push the thinking.

The tool takes the user through various steps, like

- defining the problem better, for instance looking at it end-to-end

- isolating its components, and abstracting them into more general categories
- finding relevant and interesting analogies for inspiration
- inventorying the user personas, from procurement to legal, within and outside of the company, and their struggles
- analyzing the problem through specialized lenses (such as Lean 5-whys and Fishbone, or others such as HR-specific ones)
- generating AI and other technology-based options
- scoring the results based on feasibility and novelty criteria among others

At each point, the human in the loop can redirect the machine, for instance by adding context or asking it to use different lenses borrowed from various disciplines (strategy, operations, org design, innovation).

Clear value already emerges, and this is just a start

The tool's value, in this instance, is not in having found earth-shattering new solutions to an unclear problem (though one can use such a tool for that too, and some of the solutions below were not obvious). In this case, I wanted to put myself in the shoes of the average practitioner who needs a **broad understanding of the problem and possible opportunities**, before or while engaging with domain experts. That scenario is very prevalent and immensely important because true transformation typically requires professionals from multiple disciplines to collaborate and overcome their respective knowledge gaps. The inability to understand and support each other in understanding the opportunity (the "what") and creating and assessing the desirability/viability of solutions (the "how), hinders improvements.

While the results are interesting at many levels, I consistently find this type of system particularly helpful in **exploring and structuring the problem space**. This is not something that humans typically like to spend a lot of time doing, as most professionals want to quickly move to the "solving" part of the process. However, innovation experts do know that better solutions typically stem from well- and more creatively-defined problems. (Einstein allegedly said "If I had an hour to solve a problem, I'd spend 55 minutes thinking about the problem and 5 minutes thinking about solutions" and design pros swear by "falling in love with the problem before falling in love with solutions"). A tool like this does help: it is patient, it doesn't mind doing lots of mental gymnastics, and it is pretty solid in structuring the results.

To be fair, I could have run the tool a lot harder too, even in its current incarnation.

I could have taken the machine into partial sidetracks as the tool we built easily caters to various lenses from strategy to operations, and innovation. I could have added more of my own reflections. I could have even launched an autonomous agent that looks at various lenses and then summarizes its findings for me. I could have done this with a team, during a formal workshop. Instead of 10 minutes, if this was a real project, I could have allocated a few hours - but very likely, it would have made me shrink the time-to-value by a factor of 2 or more, as it does feel like you get done in a few minutes as much as what often would take an hour, with less friction.

There are clear additional functionalities that one could build into the workflow. Things like adding more external data by feeding the system with the latest on something (e.g. new papers talking about generative AI's use cases); and adding more internal data, proprietary to the company, without complex fine-tuning (Retrieval Augmented Generation, or "RAG").

Importantly, today's models (in this case, I used Anthropic's Claude) have a good context window, which enables the tool to "keep in mind" much or all of the thread and build on it later in the process. What's really exciting, is that these attention windows will continue to grow in the future, and the workflow will be able to use previous or additional outputs even more liberally, for example by deliberately recombining some of them and exploring them as specific threads, until a final filtering and recombination. (More on this in the article here.)

Overall, the workflow shown here isn't something that any user can get out of the box, but it can be built quite inexpensively as long as we avoid some of the technical pitfalls we encountered.

The main point is: generative AI is ready to help you with in these use cases *today*. In the words of Ethan Mollick - this is already well within the "jagged frontier" - and within your organization's strike zone.

Generative AI Can Ideate Harder

The world needs more breakthroughs – climate, energy, healthcare, education, and policymaking, just to name a few - and faster.

Breakthroughs often come from the combination of ideas from very diverse origins. Humans, because of their capabilities, incentives, and the objective complexity of the current state-of-the-art, struggle with integrating that diversity of notions, at the scale required. Can Generative AI help?

In some domains, like folding new proteins, it does already – for instance, look at AlphaFold, a Google DeepMind AI program that creates new proteins. While awe-inspiring, that's possible only when AI has a good enough model of the world, can run experiments at scale, and do that largely by itself so it isn't encumbered by humans' lower processing speed. However, most AI deployed, and especially large language models (LLMs), don't know what the world is - it only knows "how the world talks about the world".

But there may be reason for optimism. Let's start with the following example.

Solving really, really hard problems

You might have heard this story, a classic when being trained on innovation design (laid out in the HBR article titled "Are We Solving the Right Problems", by Thomas Wedell-Wedellsborg): the best way to address elevator's passenger dissatisfaction at the duration of an elevator ride is not, beyond a point, to spend more on engineering; it is to put *mirrors on its walls*. To most elevator users, the few seconds lost don't matter – what matters is the *perception* of wasted time. And human perception is easier to manipulate than gravity. The result? Billions of dollars in costs saved, better safety, faster-to-realize projects.

It is also a solution that LLMs find very hard to come up with by themselves, out of the box, especially when prompted in a very narrow way ("faster elevator ride"), which is how most users prompt most of the time. Herein lays part of the solution: tell the AI to think about radically different ideas that combine solutions from different spaces, to take an expansive perspective of the user (or at least abstract away from the problem), and, importantly, to apply "lenses" to the problem. Lenses like "think of a solution as if you were the Dalai Lama".

That combination works, especially when (a) done deliberately and recursively in a chain, (b) pulling ideas from very different spaces (c) with today's increasing AI context ("memory") window, and (d) with human interventions in the right places.

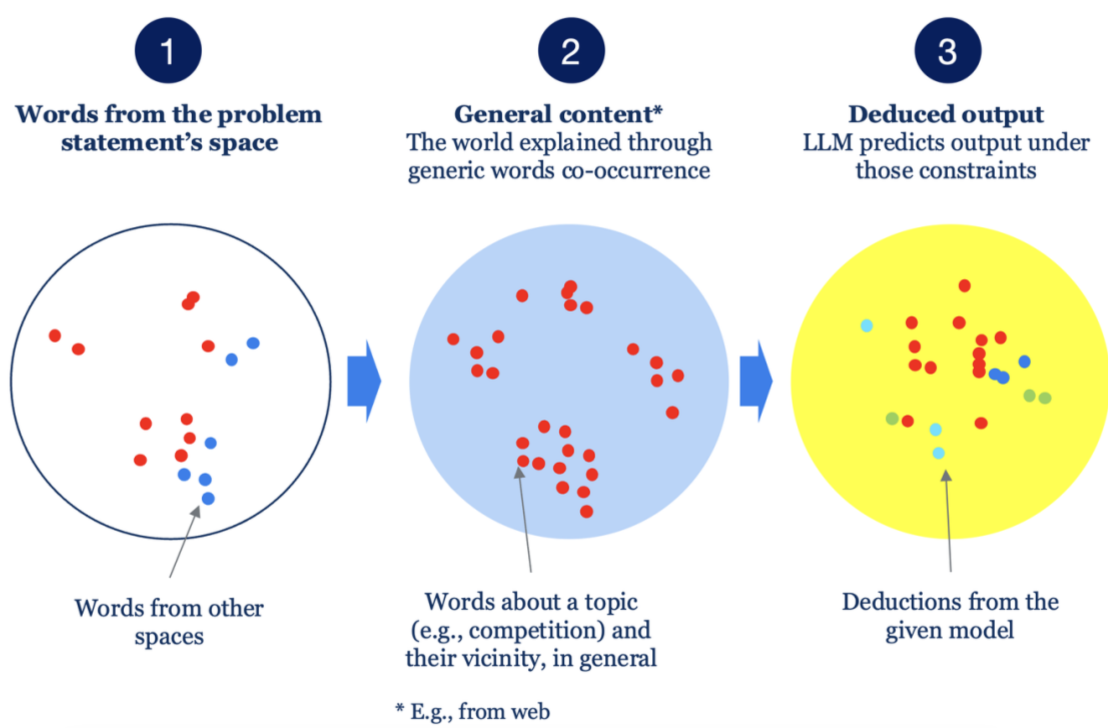
Together, (a) (b) (c) (d) may constitute a big first step in making AI more capable of creativity. Even better, this combination scales well as part of innovation workflows, as AI can sift through orders of magnitude more ideas than traditional teams could do.

Yet it is not trivial to get such an output, especially at scale. Anyone who's used ChatGPT or its predecessors knows that the output is typically well-structured, comprehensive, and convincing, but often relatively bland and unimaginative. And when you increase the so-called "temperature" (randomness, creativity), you frequently get a lot of useless hallucinations. That's at least with the versions you can easily get in your hands, off the shelf. What they do is represented below.

That's just fine for most users and uses. But for true innovation, you don't want most ideas to be "OK" – you want a few ideas to have an inordinate potential, even if at the cost of throwing away 99% of the output. How does one get that?

It is about framework-based lenses, filtering, and recombination. All the way down

Everyone, including pundits, was surprised at how much of our world's functioning is already reflected in our language, which makes LLMs so good. But truly novel ideas do not just come from the prediction of the most likely tokens based on billions of text examples.



Truly novel ideas are at least partly the *result of understanding of how the world* *works*, and the application and porting of those conceptual models to new situations.

That's where the "Dalai Lama's elevator" example is interesting. An LLM wouldn't typically go think about Buddhist wisdom when engaging in a conversation about elevators. But we – humans - could make it do it: make it look at reality, and the problem statement, through a different lens, one of attempting to influence human perception instead of blindly ramming against physics constraints) and getting it to recombine it with a very different field (mindfulness, for instance). There are two significant implications of this.

First, frameworks matter

We, as a human species and society, have *embedded* complex reasoning, and some understanding of how the world works, into artifacts that AI can mine, not just through syntax and general semantic similarity, but through *theories and frameworks*.

Lots of world-leading symbolic frameworks are embedded into the semantics of theories (e.g., "Porter's Five Forces"), authors and artists (e.g., "Andy Warhol"), or into social constructs like people's roles (e.g., "a medical doctor"). Those embed into semantics (or semiotics and visual styles, in the case of imagery) a *representation of these people's interpretation of the world*. They are the results of lengthy research processes, performed by gifted individuals and their teams, that weeded out connections between things that didn't work. In a way, they are a form of natural selection for ideas, crystallized in carefully crafted text, and propagated by thousands of examples of their use in the media, for instance. Generative AI can read that "DNA code".

We saw that through the early prompts for Midjourney or Dall-e. Things like "paint this like Andy Warhol would": for a machine, Andy Warhol *is a framework*, that Andy Warhol built with his brain and all the stimuli he processed as a person in the world of his day, and Andy Warhol's style is an embedding of his symbolic representation of the world. And we see it in the importance of "persona-based prompts" like "you are a helpful innovation consultant with experience in human-centered design, neuroscience, and construction engineering": around those words, there are many others linked to specific applications of the related methods, science, and technology. That is one set of representations of those concepts, with an explanation of how the world works as studied by design, neuroscience, and construction engineering.

Frameworks are creative constraints that have forced us for centuries to explore problems and solutions through a crystalline lens. That "passing through the narrows of the constraint" is often the spark for true innovation. And now, we can do some of that with AI-powered machines.

Second, the recombination of ideas matters

Steve Job's chief integrator role across disparate disciplines is a good example of the power of recombination: his love for calligraphy, his enhanced perception partially due to psychedelics and the resulting obsessive empathy with human reactions, his understanding and entourage of computer science, gave us computers that don't feel like computers (they feel like art). A related ecosystem of people playing with AI and Gorilla Glass resulted in "no-keyboard keyboards".

In another telling example, Wikipedia's Jimmy Wales, inspired by his knowledge of open-source software creation, applied it to knowledge curation, triggering the birth of a non-hierarchical editorial encyclopedia and revolutionizing how we look at the curation of knowledge.

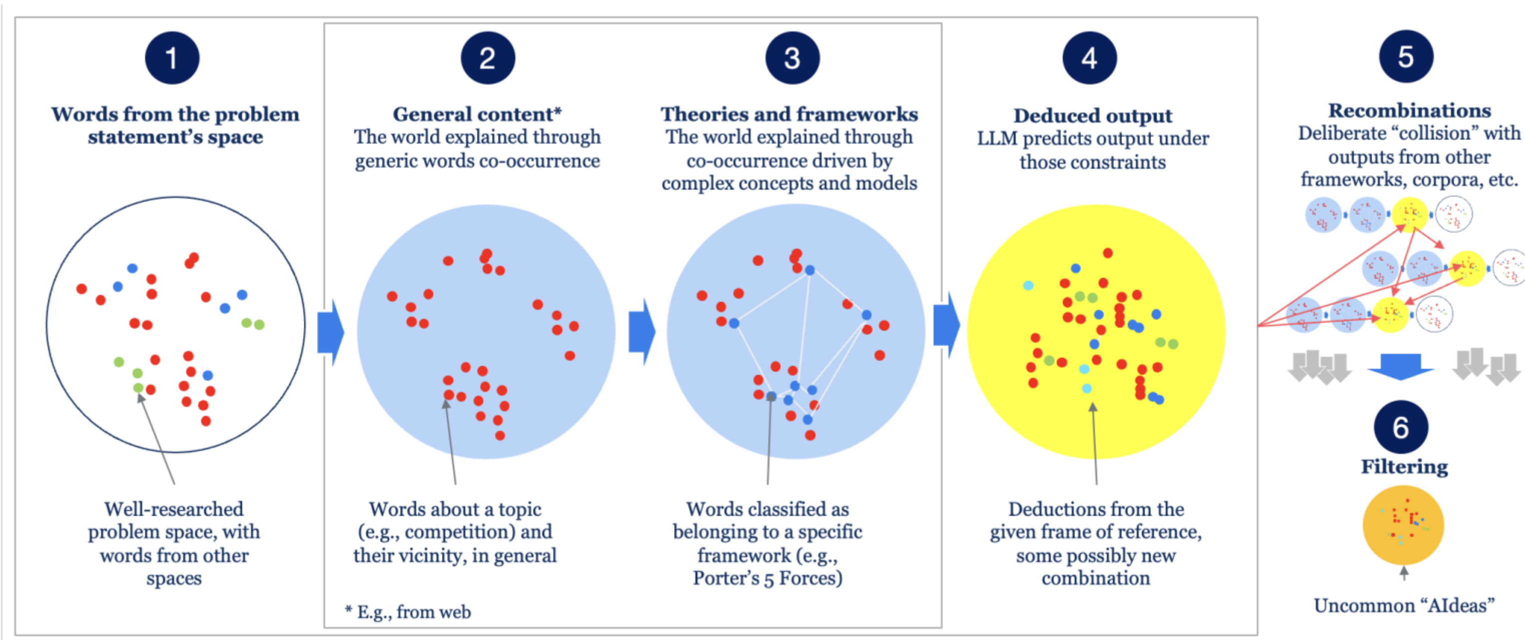
These are just two examples, of the inception of the most valuable company ever, and one of the most useful websites ever. (Apart from them, a significant body of research shows that breakthroughs in science come from the connection of ideas from different fields – among others, look up Matt Clancy's New Things Under the Sun for a thorough review, or some classics like Steven Johnson's "Where good ideas come from").

Of course, these were and are extraordinary people and teams. The good news though is that while harnessing very diverse fields is hard for a human, for AI the distance between knowledge items is a computationally tractable problem. Machines don't suffer from the so-called "burden of knowledge" which limits the rate at which people can achieve enough competence to be able to contribute something novel in their field, the same way as we do. They don't have "working memory" limitations the way we do.

Theoretically, generative AI tools have most of what they need already - they just need to be pointed at it.

Enter "AIdea Colliders"

What we need is something like what is described in the following chart. I tentatively call it Deliberate Framing, Recombination, and filtering (DFRF). It is, in other words, an "Idea Collider", loosely inspired by the work of colliders in physics. The output is "*AIdeas*" (not a typo), triggered by the AI's language-reasoning capabilities and the deep knowledge embedded in human theories and frameworks. In an AIdea Collider using DFRF, problems are (1) exploded and explored, and LLMs outputs are constrained through theories and frameworks (2) and (3), before (4) the deduced AIdea outputs are recombined (5), and then filtered (6), at scale.



An example of this architecture is **MIT Ideator** from the Massachusetts Institute of Technology's Center for Collective Intelligence, an idea-generating machine built by our lab's team after GPT-3 was released. There, we do away with the traditional chat interface and force the machine to apply a series of lenses to the problem statements, then ask humans to recombine the pieces. The focus of the tool is to help solve problems through a specific design framework called Supermind Design. It has a strong emphasis on exploration of the problem space, and on organizational designs that leverage collective intelligence. A full paper is being released on Arxiv, and you can apply to use MIT Ideator and give feedback.

Think [encyclopedia]+[web]+[community]=[Wikipedia]. You would know the answer today, but not before Wikipedia existed. Many other examples exist (you can find some on the supermind.design website), and countless more could be built.

More broadly, idea colliders can be built with **frameworks from a very wide range of spaces.** Consider some illustrative examples:

- **strategy** e.g., Christensen's disruption, Blue Ocean, PESTEL, Experience Curves, SWOT
- **innovation** ideation e.g., Design Thinking activities such as journey mapping, persona analysis, or analogies (also called alternate

worlds), Doblin's Ten Types of Innovation, TRIZ, Lean Startup, Six Hats

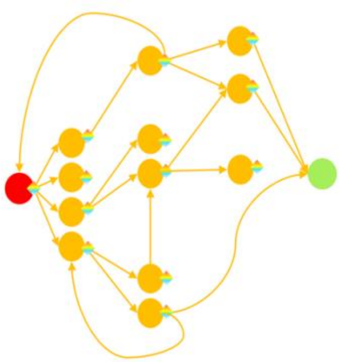
- **decision making** e.g., cynefin, logic tree, Eisenhower Matrix, balanced scorecard, debate techniques
- **operational improvement** e.g., Lean Management's FMEA or RCA, Six Sigma practices, ISO 9001, HAZOP
- any **other management framework** e.g., McKinsey 7s, Ray Dalio's Principles, Ikigai, Agile, Theory X and Theory Y, OODA loops, structured coaching methods, Peter Drucker's theories and principles
- industry-specific **frameworks** e.g., Consumer Products HACCP - Hazard Analysis Critical Control Points
- and **many others**, including art (e.g., Brian Eno's Oblique Strategies traditionally used for music, coincidentally also highlighted by Ethan Mollick recently), psychology (e.g., personality types, Maslow's pyramid, flow theory), ESG parameters, personal coaching (e.g., Ikigai), or any logical and structured reasoning method (e.g., induction/deduction, Socratic questioning), etc.

And AI output can be fed with interesting, specific examples data sourced from diverse and faraway spaces, for instance, embedded in vector databases for retrieval-augmented generation. Think about being able to mine Arxiv, Patent Office records, Crunchbase startup databases, healthcare guidelines, or climate mitigation practices (including those successfully used in developing countries, or by indigenous groups) among others.

The **recombination and filtering** phases will also be crucial. An AIdea Collider can theoretically generate millions of idea fragments – and recombine them, geometrically expanding the output. Some of the triaging can be done by machines, for instance through specialized models representing a digital twin for the typical desirability/feasibility/viability scorecard process; or with composite machines based on an ensemble of models, able to critique each other – for instance, a model that helps fact-checking or identifies similar solutions, or one with stronger ethical skills applied to the output.

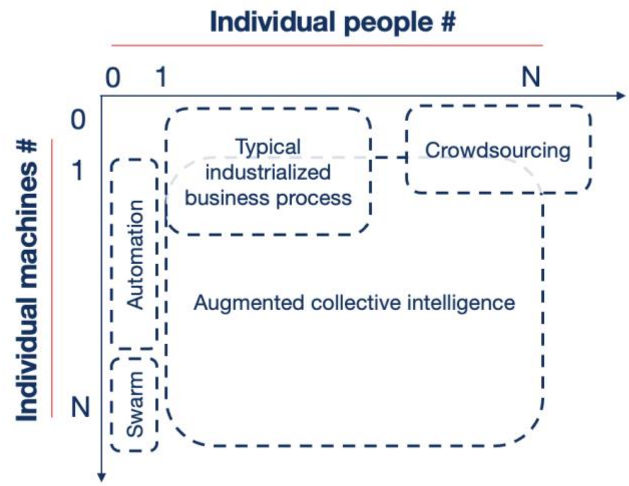
Part of this work could be done by humans, especially in large networks to distribute the load and harness varied viewpoints. People could "prune" specific branches of the output and give more emphasis to others, for instance with AI providing summaries and mapping the exploration space more visually, like a hyper-scale form of sticky-note clustering that innovation professionals are familiar with. At any rate, it is very likely that, to generate real breakthroughs, AIdeas will often require augmentation and recombination by human experts who possess innovation and domain skills.

Ultimately, we need for thinking of ideation as a scalable process, with many steps, some of which are recursive, as illustratively depicted below.



These steps, like those in any other deliberately engineered business process, will feature a *combination* of machines and people - what is called a *supermind*. Not unlike what one would do with people, some of the machines will have the same skills and will possibly only use different lenses, while others will be different (as in the example of the ethical AI filter mentioned before). The space for designing such processes is much vaster than what we traditionally use, as visualized below.

Remember the world before the internet, automation, and Wikipedia? Compared to today, the future will be as different as today is compared to that long-gone time.



Start small and fast. But also, this is the time to think big

To recap, the basic workstreams for an Idea Collider are:

1. Identify relevant **frameworks** and classify them
2. Identify interesting **data** sets
3. Create chained **flows**, with the ability for humans to prune intermediate results, and recombine others
4. Work on the **right UI/X** to enable frictionless human-machine collaboration
5. Work on models for **filtering**

This is just with current technology, and a minimal amount of coding. There's a lot more, but this already can get many started to build a real, large-scale Collider, with evolving contemporary AI capabilities such as chaining, context windows, multiple agents (and potentially even swarms).

The future possibilities are manifold. For instance, AI models increasingly use tools, such as OpenAI Plug-Ins, to perform specific tasks, and Colliders could be a new type of plug-in. Individual organizations could build their own – for instance, consulting firms and innovation departments - provided they allocate the right amount of data engineering capacity and design it well.

And that's even before deliberately mining knowledge graphs, hence adding a layer of signals that maps connections between topics, between people, and between people *and* topics – ultimately yielding additional ways of exploring the solution space.

All of this is ready for productive experimentation for quite a few use cases. Innovation organizations that use AI-augmented collective intelligence (ACI), for example in the form of Idea Colliders, stand a real chance to accelerate the future by unlocking the breakthroughs we all need.

Start today.

Your Problem-Solving Idea Flow, AI-Augmented

How do we use AI to augment human capabilities to generate better ideas and solve problems? There's immense potential in doing this well, both personally and organizationally.

We talked elsewhere about recent research that could be of practical, immediate use to professionals: the general vision of augmented human capabilities and how we need to lead machines to ask us questions; how we can get AI to ideate harder; the work that we have done at MIT to provide a scaffold, an exoskeleton to humans in the ideation process; the competencies that human professionals need to acquire and some relevant skills to wield the power of AI better. These contributions are worth reading as they complement what we will do in this essay.

Here, **I share practical guidance that your teams (and their machines) can execute to generate, day after day, better ideas**, whether it's problem-solving (known-knowns) or creativity (all the way to unknown-unknowns). I will detail some steps to incorporate AI into current work practices, leading their people to use these steps as a standalone method or as part of their existing techniques. As you explore these concepts, remember that while I mainly refer to text-based problem-solving and creativity tasks, AI can increasingly be used for visual (and possibly spatial) and even auditory tasks.

A few reminders first:

1. **The process is not one of querying machines; instead, it is about *augmented collective intelligence*.** There's a real risk that humans might let poorly managed or incompetent machines take over, leading to mediocre or worse results. This is, unfortunately, a common issue in many organizations, and many of the disappointed accounts we hear in the media can be traced back to this problem. Similarly, there's a misconception that "AI knows everything at any one time, so there's no reason to tease its thinking through a process. As the evolution of prompting techniques demonstrates, there is great value in leading machines' thought sequences. So, put your pilot gear on and take control.
2. **Use individuals, groups and machines in deliberate sequences.** The general instinct is to have individuals query individual machines. I recommend considering different configurations, such as: (a) humans think first individually, then in group(s), and then add the AI's perspective to theirs; (b) the perspective of humans is injected into AI for the machine to build on it (c) the perspective of machines is injected

into human groups for people to work on them; (d) all of the above can be done with multiple AIs instead of one, for instance, agents that have specific roles (planner, curator, critiquer, etc.) and capabilities (broader generalist or more specialized models; multimodal or not; etc.).

3. **A problem-solving or ideation process is *not* linear**. The gestation of ideas, especially great ideas, often involves a non-linear, meandering journey with recursion and back-and-forth. This is a normal part of the creative process and should be embraced.

We need frameworks for multi-step AI augmentation. Some literature on this is emerging (like our MIT and some recent Harvard work), showing that well-designed human scaffoldings improve the quality of solutions. These early experiments hint at the possibility of multi-step flows and their potential. Here, we dive deep into what those steps can be.

The overall structure of the flow (below) will be familiar to many, especially those familiar with sophisticated problem-solving or ideation processes, such as design thinking, especially in its double-diamond form. This structure will guide us through generating better ideas and problem-solving with the help of AI.

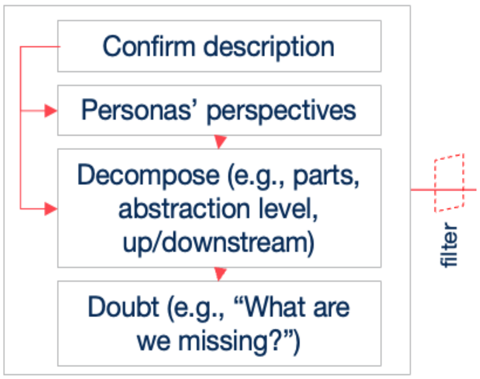
First Step. Exploring the Why: AI Helps

Falling in Love with the Problem

The first chunk of the process consists of exploring the problem space. You might have heard the phrase, “To generate better solutions, you must fall in love with the problem before falling in love with the solution”. You might also have heard people saying that it is essential to define the Why of the problem first and the What (category of problems it belongs to). Not spending time understanding why a problem matters in detail or segmenting the problem space improperly typically leads to poor results. The good news is that AI can help us do that; the bad news is that humans don't automatically do that with AI, yet. The additional value of exploring the problem well is that, apart from helping humans, it helps the machine spend computing resources accumulating and focusing on relevant concepts that it will later "keep in mind" for the solution-finding phases. Locking in on relevant context is essential for machines to do their job well.

Below, I'll walk you through the steps, using a specific example of a complex problem (reskilling for AI) to show more clearly the applicability of these concepts. I will first provide the schematic representation for each phase and then describe it in detail.

I recommend the following steps when it comes to discovering the Why:



1. **Confirm problem's description.** Enlist the machine's help to confirm the problem's description, hence trying to avoid possible "loss in translation." Language is not a very precise tool at times, but language models, if guided appropriately, can help us refine it. For instance, if your problem is one of "finding creative ways of retraining executives so that they harness the power of AI," you should lead GenAI to ask you and your team additional questions about what you mean by that - which could result, for example, in spelling out what kind of company you're working, what kind of people work there, which people are most at risk, what kind of resources you have already applied to this problem, etc. This is the right time to add any insightful perspective on the problem that you already have prepared - don't expect that AI "knows everything": it probably does know much of what you would know, but for cost efficiency reasons, it is unlikely to prioritize all of the relevant information during inference time (when you query it). One interesting twist is that the recently released OpenAI o1 models, because of their ability to think logically, could significantly improve the already impressive capabilities the machines have in this space.

2. **Personas' perspectives**. Then, if the problem lends itself to it, ask the machine to map the stakeholders and take the perspective of different personas involved in or experiencing the problem we are trying to solve. In our example, that might mean different seniority levels of workers (for example, CxOs or entry-level employees), in various functions (for instance, in the finance or the sales function), in different parts of the world, and needing different things (for example, a general overview as opposed to practical knowledge required for upcoming business meetings). Use AI to describe the journeys of each of those personas as they traverse the situations that cause problems, and ask AI to journal

their emotions (e.g., an employee trying to find time to learn during a busy week and then struggling with identifying the most relevant and appropriately-sized piece of content for that specific day). Or, if you have already obtained traditional client or customer preference data, include it as context.

3. **Decompose**. Third, apply some decomposition of the problem, for example, breaking the problem down into parts and types, identifying the precursors and the consequences of that problem, as well as moving up and down the abstraction ladder to identify levels of specificity or generality that could yield a better vantage point on the problem. For instance, here, learning new skills for AI means a set of specific learning steps but also includes upstream problems like the curation of contextual practical knowledge that can help adult learners understand the problem in the context of their work and downstream issues like the inability to teach subordinates and colleagues what one has learned. Abstractions are essential when solving seemingly intractable problems - since, in this case, AI could help us understand the broader set of problems related to ours, for instance, general knowledge management issues or knowledge transfer challenges present in many new-technology innovation periods. Again, new, more logically-reasoning AI models like OpenAI o1 can help here.

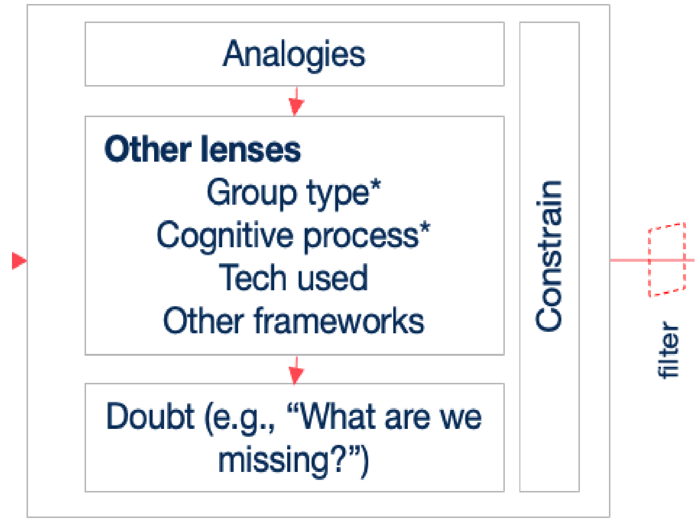
4. **Inject doubt.** Enlist the AI tool to poke holes in our description of the problem. Generative AI is often quite good at critiquing thought processes, especially in textual form, and could help identify gaps or white spaces that could be meaningfully explored before moving on to the next steps. You can also ask AI to search for knowledge on important dimensions of the problem to bring some more insightful angles of attack to the fore.

5. **Summarize and filter**. After this, it is essential to take stock of what has been covered and cluster and further structure ideas to streamline your output, consolidate your understanding, and weed out duplications. This is also when you - individually or as a team - filter perspectives that don't seem promising (insightful, practical). Identifying what’s particularly insightful helps prioritize aspects to be kept front

and center of the AI memory - which, while being increasingly vast, benefits from being shown what good insight is. AI is increasingly good at helping filtering, especially if you clearly define what is important for you (e.g., only surprising insights or MECE - mutually exclusively and collectively exhaustive ideas, etc.). This is also an area where new, strong logic models can help.

Second Step. Exploring the What: Guiding AI to Interesting Solution Spaces

The second part relates to discovering the What, that is, understanding what category of problems and solutions are relevant. I like to do a thorough job at this often-skipped phase, and people move directly to generate ideas. From experience, subtle yet powerful insights can be generated by looking at the problem through a creative, unusual, categorical lens. As an example, there is an old innovation saying - "if you are segmenting your market the wrong way (typically, the traditional way", you will never be able to truly innovate" - which illustrate the value of taking the right type of perspective. There are many ways of doing that, and AI is very useful here.



1. **Analogize**. Start by finding insightful analogies. AI has become better at this, especially when given enough context. In our example of up/reskilling for AI, analogies are: individual plants in a forest that become aware of and react to the threat of a new parasite; or community workers in rural villages of developing economies needing to learn how to administer vaccines. Good analogies often shed light on poorly understood parts of the problem (for instance, the behavior of decentralized organizational units in our example). They can yield interesting avenues for solving those problems cost-effectively (for instance, in our example, by hinting at the energy communities of employees can muster).

2. **Use lenses**. Then, apply other lenses that force your AI to think of

the problem in new ways. At the MIT Center for Collective Intelligence, we developed some of them, dealing with the design of organizations, as part of the so-called "supermind design methodology". For instance, what kind of organizational structure ("*group type*") is currently trying to solve the problem? (In our example, most learning is managed hierarchically, with some opportunities for communities of interest to form.) Another lens is the *supermind's cognitive process*, which examines what the collective brain of the current organizational design is trying to do. (In our example, the whole organization tries to learn by creating capacity in the appropriate parts of the collective brain, while specific units are "sensing" the environment to identify the correct use cases people should learn about.) Similarly, understanding what technologies are being used and their advantages and shortcomings is helpful for the subsequent solution-seeking phase. Finally, I can't emphasize enough the value of encouraging AI to apply to the problem existing, human-made relevant management or scientific (or other) *frameworks*, to nudge AI into emulating some symbolic/abstract thinking embedded into the vast literature that exists around those frameworks - e.g., in our example, using Christensen's Disruption Theory to unearth interesting dynamics relative to skill-based competition (for instance, AI "disrupting from below" the work that humans do, by taking first mundane and repetitive tasks, with human incumbents doubling down on their efforts to perfect their current capabilities instead of migrating to a different value proposition while they have enough resources).

3. **Inject doubt**. As in the previous phase, leverage AI to poke holes and identify gaps in this phase's analysis. "What are we missing?" is an excellent question for your AI and yourself.

4. **Work with constraints**. Truly creative thoughts may emerge by forcing the thought process to navigate hard constraints. For instance, in our example, you could use AI to detail extreme scenarios, such as those where learning resources are extraordinarily scarce or abundant, where the learners are highly sophisticated or extremely junior, or where all learning opportunities happen asynchronously and remotely, as

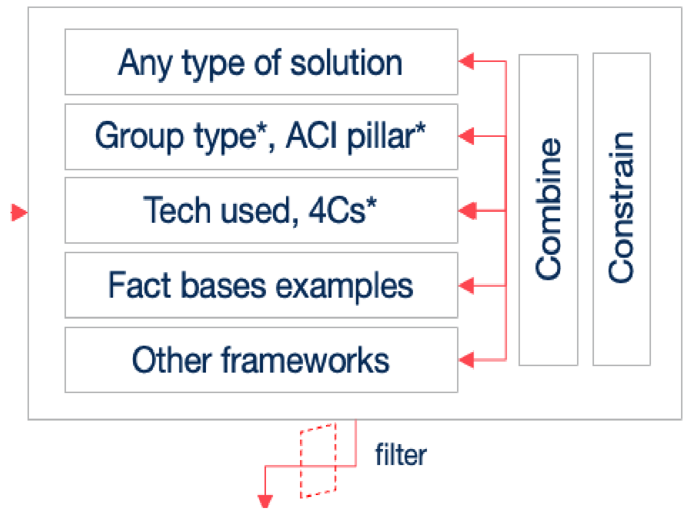
opposed to others where all learning happens in person with real instructors. Constraints also come in the form of customer preferences (say, durability vs. performance), industrial constraints (e.g., materials availability, production location), or budget boundaries (e.g., limited or conditional budget).

5. **Summarize and filter**. After this, as in the previous step, AI will be leveraged to summarize and cluster ideas and filter out less-interesting thoughts with your and your team's active participation.

Third Step. Exploring the How: Co-creating with Machines

Now that you've "fallen in love with the problem," it is a good time to venture into solution-finding. This last phase relates to activities that help generate solutions using AI as a thought partner. That means avoiding using it as an "oracle" and instead ensuring that you and your team are active participants in iterating ideas as they form. Those who resist the urge to move too fast into this phase typically deliver more novel and effective ideas.

You can start by asking AI to re-read the individual Why and What threads as context and deliberately identify possible solutions. A few good ideas might emerge even at this stage if the problem exploration was sufficiently insightful, for instance, through truly eye-opening analogies. After producing the first ideas, do not hesitate to point AI's attention to the output of specific previous steps (e.g., "What does the analogy XYZ make you think?"). Ensure you don't anchor the following steps on these first results to



avoid missing genuinely creative ideas.

1. As in the previous phase, one should force AI to generate solutions through lenses derived from frameworks and management (or other) theories, such as the following.

- **Groups**: The supermind design's "group type" described above is one (for instance, highlighting and facilitating the teaching role of practice leaders as part of communities of interest).
- **ACI pillar**s. The Augmented Collective Intelligence Pillars (more at Supermind. design) is

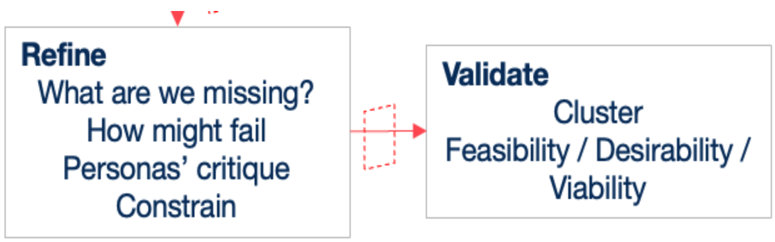
another, leading AI to identify ideas that solve for specific high leverage points in collective-intelligence systems design (in our example of upskilling at scale, by for instance, enabling an AI-assisted, decentralized curation-at-scale required to surface relevant, contextual learning examples; or identifying the most important network nodes, humans and machines, in a network and enabling them accordingly; or creating the right incentives for a community; or providing a proper collaboration infrastructure).

- **Tech use.** One can also ask AI to provide ideas about using specific technologies (say, augmented reality or edge computing) and apply them to parts of the Why and What components. This is the area that most people intuitively think about first. Still, technology adoption is most effective when tied to organizational design components, which means using it on the right other "how" levers.
- **Fact bases.** If a database of solutions examples is available, with highly curated and relevant data sets, it could also offer valuable stimuli for AI (for instance, a Tech Crunch database of all education startups or a database of collective-intelligence organizations such as the one available on the Supermind.design website).
- **Recombine**. Most ideas are of limited interest, but their recombination could yield surprising value. AI can help recombine ideas at scale. They can converge into alternative solutions described as one would typically do with so-called "concept posters." Models that are better at reasoning, like OpenAI o1, can complement the creativity of older ones by combining and deduplicating ideas, etc.
- **Constrain**. Once again, have AI develop ideas under specific constraints, such as available resources and capabilities or specific customer requirements and preferences.
- **Filter ideas once more**. You want to encourage the AI to over-emphasize specific types of ideas. For example, privilege the most creative or surprising or use the most understandable and potentially useful. Keep in mind that many models will try to satisfy you by offering ideas that do not depart too much from normalcy, which means that you and your team must be prepared to forcefully prod AI to go into uncharted territory (or you might need to fiddle with system

parameters such as model output temperature).

Once you have filtered, with the help of AI, the numerous inputs from this phase, you are ready to **refine** them. Several activities could get you there with the help of AI tools. I want to underscore how valuable AI can be in this phase: humans are often not great at candidly critiquing ideas, as they're bounded by what they know and want to respect civility conventions - but AI can find holes into seemingly good ideas, if it is asked to look into logical cracks, play contrarian or antagonize, among others. As many innovation experts say, "Ideas are cheap, but finding flaws in them is very hard and valuable."

1. First, once again, use AI to identify **blind spots**. Has your process inadvertently missed out on a specific range of potential solutions? Are there implicit biases? For instance, in our example, has AI suggested solutions that are out of reach for older employees?
2. Then, analyze how each idea or component of ideas could **fail** to meet its objective. AI can support a thorough exploration of the failure modes. For instance, in our example, AI could highlight managers' reluctance to give their teams guidance and time, limiting the uptake of any reskilling program.
3. You can also take the perspective of **different personas** and **critique** each idea from their viewpoint. For example, what would the CFO think of the ideas? What would a junior employee in South America think of it (e.g., "Are you providing self-improvement coaching in Spanish?"), as opposed to a senior leader in Japan (e.g., "How do you provide constructive educational feedback in group forums while showing respect and deference to authority?")
4. Finally, use some **constraints** again, but in this case, their value is to weed out potential solutions, for instance, those that require too many resources or have too long of a deployment cycle.
5. The last filtering exercise can be **clustering and structuring** the remaining ideas and asking AI to assess their feasibility, desirability, and economic (or other) viability. AI can do that in a range of reasonably well-understood domains and could help even in more specialized domains by using first-principle thinking, especially if given access to a range of successful or unsuccessful examples from existing databases.



Enlisting the Full Power of Humans and AI to Push Hard

The process that we just described can be particularly successful when leveraging a few opportunities:

1. **Human in the loop**. Tools can create many ideas quickly, but designing a human-centered experience and training your users to remain firmly in the loop is essential. That means providing continuous and intellectually aggressive feedback to the machine to avoid drifting towards platitudes. You can also ask part of your team to do the same exercise as AI does, in parallel, and then converge the results. Or you do the exercises first, with AI asking you probing questions.

2. **Using different agents**. It is not hard to build AI agents with specific capabilities that you can use for specific steps. For instance, an AI agent could be configured to be a particularly insightful "gap finder" by giving specific examples of how to do so. Other agents could be beneficial in understanding how to apply management framework, including finding the relevant ones. Some agents could be configured to be good raters of possible solutions. AI agents could credibly take the perspective of individual personas.

3. **Asynchronous batch processing** for "fractalization." Conversely, as AI is increasingly cost-effective, you can build workflows that potentially follow many "rabbit trails," branch out from those, and provide intermediate and final results at specified intervals. For example, some of these machines could use brute force to combine many ideas, e.g., trying to combine every group structure and technology type. This only works if your filtering mechanisms are reliable, but that should be the case in specific domains, especially over time.

4. **Recursive loops**. As I mentioned at the very beginning of this essay, do not mistake this process as a step-by-step, one-directional checklist. Solid problem-solving and creativity require the ability to trace back multiple steps, start again, deemphasize specific trajectories, and try initially neglected ones. And that is, for now, firmly the work of a human.

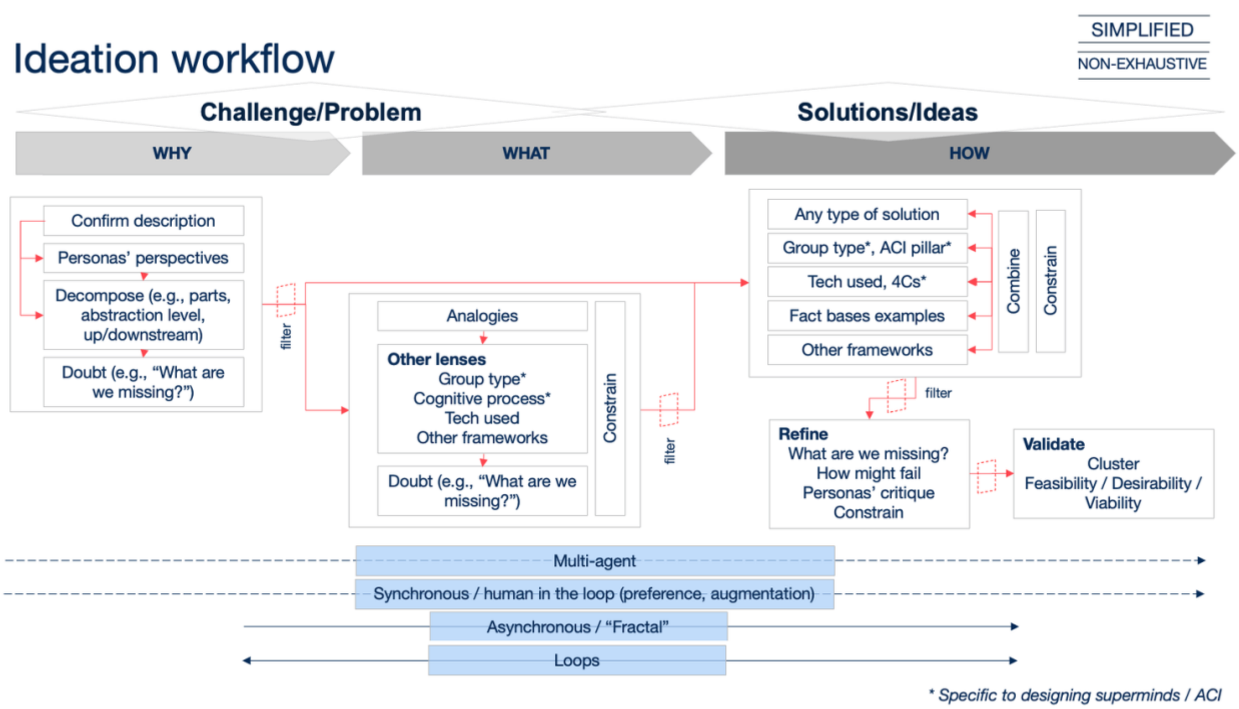
Below is an overview of the entire process we discussed. Remember that this picture is not exhaustive: other problem-solving and creativity techniques, particularly those that you and your teams are used to, could be inserted as part of the overall process. That's especially powerful if they're framed correctly and accompanied with sufficient context so that AI understands them and if you see them as a two-way street facilitating a dialectic between humans and AI, not a "download" from the machine onto us.

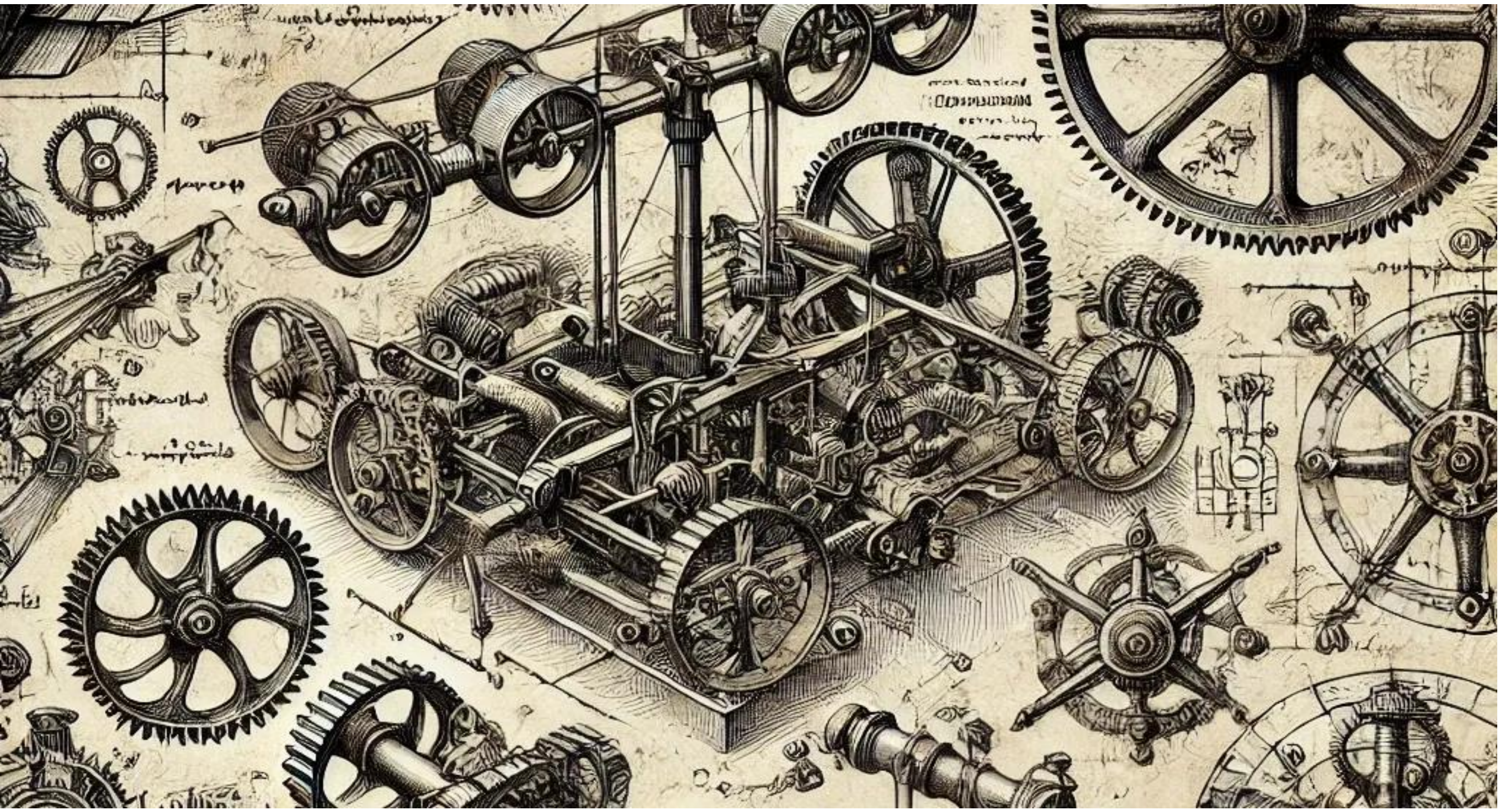
This won't stop. Get Ready for the Change

We will continue to discover how to make our tools more effective at interpreting each step. It is also possible that we will find different pathways for creativity - the idea of fractal exploration mentioned before, coupled with the usage of pre-existing mental frameworks, especially when harnessing AI to do filtering at scale, hints at possibilities that were inconceivable when using only humans as part of the creation process. Some form of truly "alien" intelligence could help us in ways we don't fully imagine today.

Time will tell, and we will discover pitfalls above and beyond the potential threats we intuit right now (such as an exponential increase of mediocre output, the increasingly ineffective use of truly bright, creative human minds, etc.).

For now, however, it seems sensible to continue the discovery of AI-assisted problem-solving and creativity, especially by exploring the synergy between humans' capabilities, individually and in groups, and the capabilities of one or many machines.





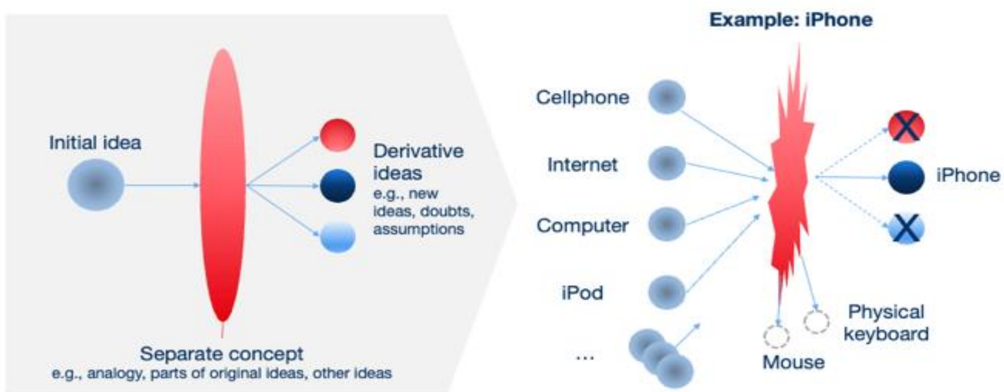
Ideas "Physics and Chemistry" with GenAI

We have discussed how AI can help problem-solving and idea generation, mainly by boosting unexpected collisions between concepts. In this article, I lay out some notions leading to building better workflows and organizational structures that natively harvest the emergent opportunities in our problem-solving and innovation ecosystems.

The Structure of an Idea and the Power of Collisions

This will not be a theoretical discussion. We know well that our ideas - our organization's, our team's, our own - often simply result from interaction with other ideas through connecting with people, reading about something new, etc. The flow of our days is a constant opportunity to generate new ideas, not because we think about things harder but because we *bump into things that trigger them*. Leaders try to foster those collisions, including through organizational and process design or encouraging the formation of active **knowledge networks**. Those collisions happen daily, at scale, and intelligence emerges from those interactions.

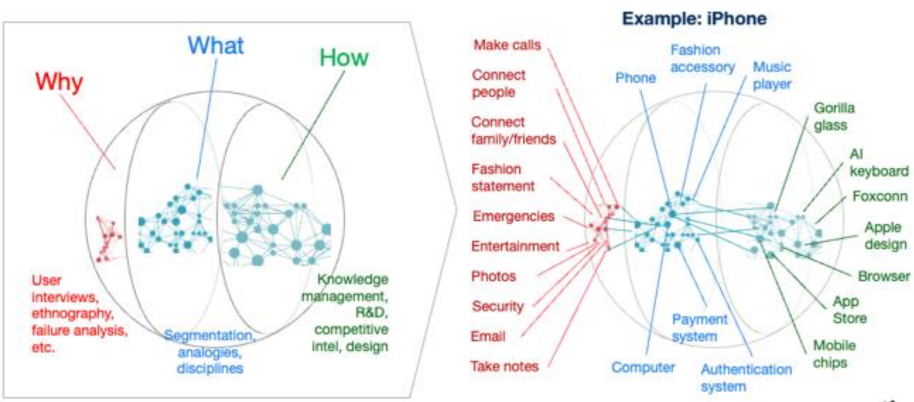
When ideas collide—whether with other ideas, analogies, or subcomponents of the original idea—they generate derivative concepts. These derivatives can take many forms: new ideas directed at solving the initially stated problem, clarifying assumptions, or even clarifying doubts that spark further inquiry. Understanding this process is key to leveraging AI for idea generation and innovation.



Let's take the iPhone as a simple, well-known example. Its creation wasn't just the merging of existing technologies like cell phones, iPods, or computers. Instead, the components collided and evolved, yielding the iPhone and countless other ideas. For instance, the removal of the physical keyboard and mouse marked significant shifts in user interaction design. This demonstrates that ideas, when recombined, are not just additive—they can transform into something entirely novel.

At the heart of this process lies the structure of an idea, which I like dissecting into three components: **the why**, **the what**, and **the how.** This is a version of a process called morphological analysis, which uses many different parameters - but I find these three to be both insightful and practical for our purpose here.

For illustration, imagine an idea as a molecule comprising atomic subcomponent ideas. The subcomponents of these three (why/what/how) components are connected and relate to each other, which could be formally represented as a knowledge graph.



1. The Why: Purpose and Importance

The "why" represents the motivations and values driving an idea. It explains why the idea matters and what problem it solves.

The iPhone's "why" included enabling communication between family and friends or others, making a fashion statement, and acting as a tool for emergencies, entertainment, and productivity.

The "why" helps uncover the emotional and functional needs an idea fulfills, often through tools like user interviews, ethnography, and failure analysis.

2. The What: Categorization and Context

The "what" defines the problem categories an idea addresses. It connects the motivations of the "why" to broader contexts.

For the iPhone, the "what" included overlapping categories like phones for communication, music players for entertainment, computers for productivity, and payment and authentication systems for security.

The "what" helps locate an idea within an ecosystem, revealing its connections to adjacent fields.

3. The How: Feasibility and Execution

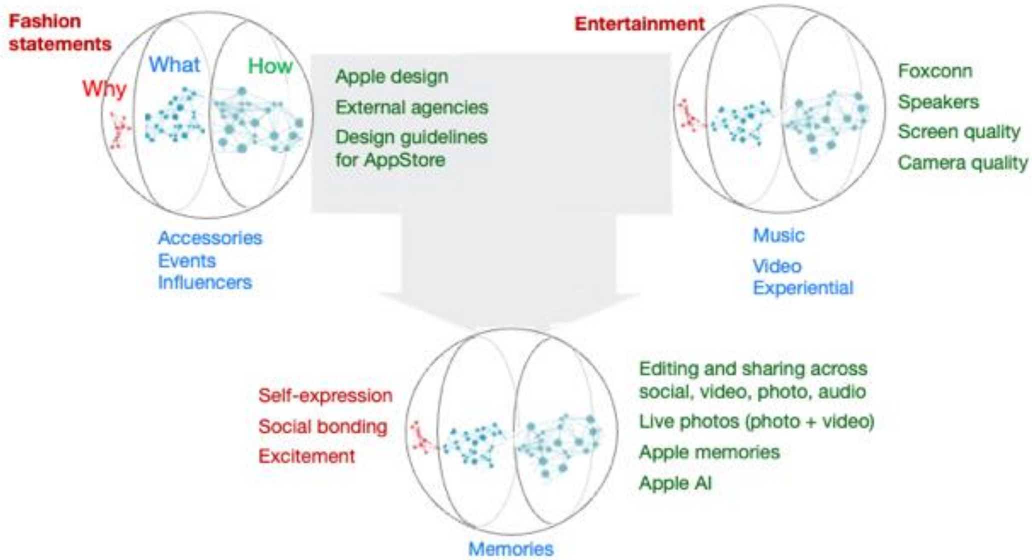
The "how" encompasses the technologies and systems that bring an idea to life.

For the iPhone, the "how" included durable materials like Gorilla Glass, AI-powered on-screen keyboards, Foxconn's scalable manufacturing ecosystem, Apple's design philosophy, and the App Store infrastructure.

By bridging the "what" and "why," the "how" transforms concepts into actionable solutions.

When Ideas Collide

Idea collisions create fertile ground for new concepts. Consider two subcomponents of the iPhone: fashion statements and entertainment.



Fashion Statements:

Why: People seek self-expression and social recognition.

What: Accessories, influencers, events.

How: Apple's design language, external agencies, and the design guidelines for the Apple Store and App Store.

Entertainment:

Why: People desire immersive, engaging experiences.

What: Music, video, experiential entertainment.

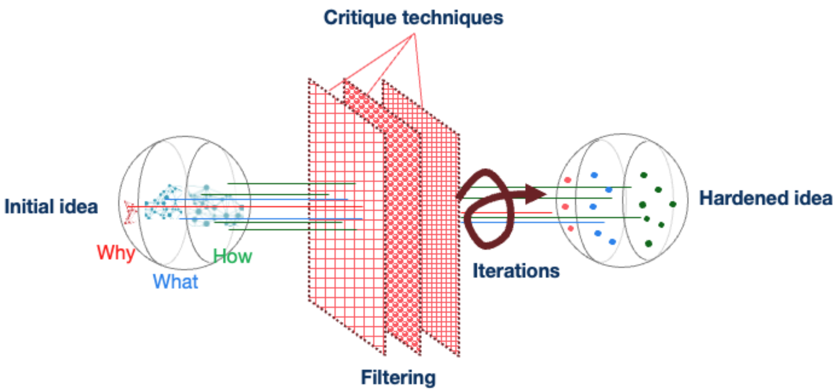
How: Engineering components like mics, speakers, high-quality cameras, and screens.

When these two collide, derivative ideas emerge. For example, combining fashion and entertainment creates intersections like shared memories, self-expression, and social bonding. Apple's "Memories" feature exemplifies this, using AI to surface curated photo and video highlights connected to people, places, and moments.

Collisions don't just help with the divergent part of idea development - they also help *refine* ideas. A special type of idea collision is "**idea hardening**": Every idea, no matter how groundbreaking, begins fragile and imperfect. For it to become robust and actionable, it must undergo a process of hardening—a deliberate refinement in which assumptions are tested, weaknesses are exposed, and components are iteratively improved.

Hardening isn't just about identifying flaws—it's about strengthening an idea's *why*, *what*, and *how*, ensuring it has the clarity, feasibility, and resilience needed to succeed.

This hardening process often relies on **critique techniques** such as persona-based evaluations, failure analysis, or "what-if" scenarios. These methods simulate how an idea might perform under various conditions or perspectives. Through iterations, feedback loops, and testing, ideas evolve into mature solutions that can withstand real-world challenges.



And crucially, the techniques used for hardening are special lenses and embed the knowledge of the people who have created them. In a way, they force ideas through some contact with specific realities - many of them have been constructed by humans.

Ecosystems of Ideas and the Role of AI

The world's idea-generation process doesn't happen in isolation, in a purpose-built vat. Humans engage in "natural experiments" daily, generating and refining ideas through interactions with people, environments, and, now, machines.

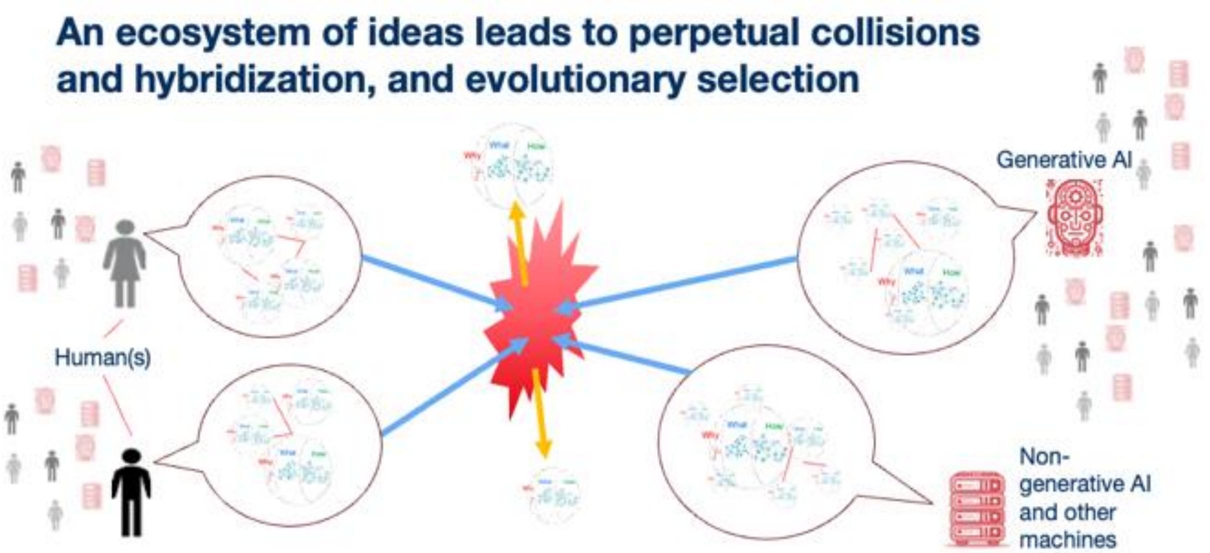
The scale of this is **massive**. We are talking about trillions of interactions daily, an unstoppable chain of collisions whose outcome, with some notable exceptions, is so far captured manually or in barely digitized workflows and processes. AI of many types changes that landscape, as AI can now access publicly available data and, increasingly, organization-specific knowledge. It can use the signal from knowledge graphs showing the relationships between ideas and the one from network structure analysis (who or what "says" what, how they're connected, how central they are in the network, etc.).

AI, with its capacity for interpolation and abstraction, enhances this process, complementing human (and human systems') extrapolation and **serendipity**.

Interpolation: AI can analyze vast data sets across diverse dimensions, generating connections humans might miss.

Abstraction: AI generalizes concepts to uncover solutions transferable across domains.

This interplay between human creativity and AI's computational power expands the innovation landscape.



Idea Physics and Chemistry: a New Discipline?

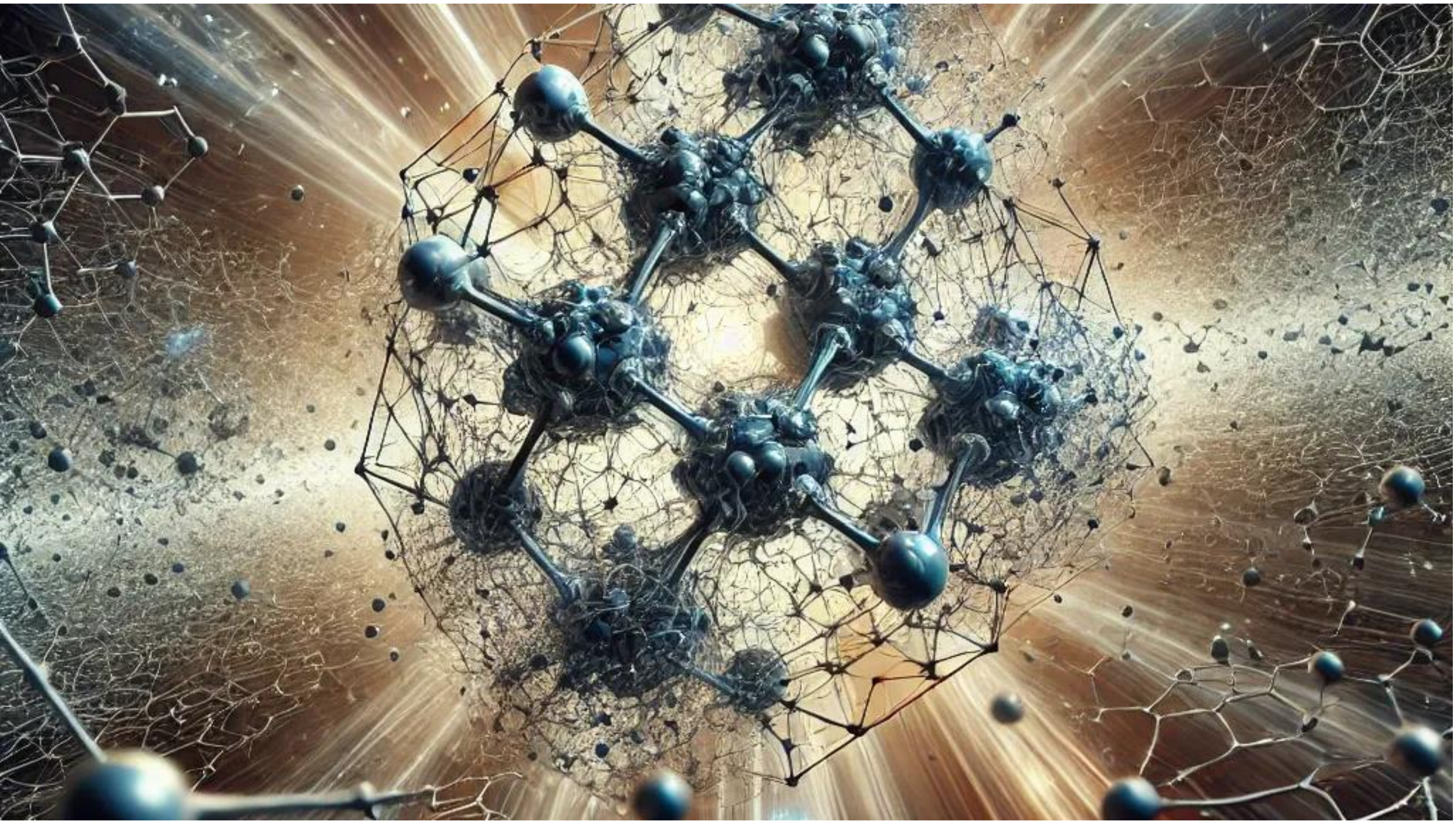
The "idea physics and chemistry" concept, a sort of combinatorics, envisions ecosystems where ideas collide and evolve, supported by AI. It is not a theoretical construct. I see it as a set of principles and frameworks driving the design of organizations, for instance, maximizing their "collision surface" and enabling their people to process that emergent knowledge.

By harnessing these techniques and tools, we could even create "chain reactions" of innovation, exploring (not just generating) new ideas at an unprecedented scale.

Imagine a world where AI helps us harvest ecosystems of ideas, optimizing the way knowledge flows and transforms within or outside of organizations. This supermind—a networked intelligence emerging from human-machine interactions—could hugely benefit how we innovate, solve problems, and create value.

By understanding and leveraging the anatomy of ideas and the power of collisions, we unlock the potential for continuous, scalable innovation. The future of ideas lies in how we connect and recombine the "why," "what," and "how" with the tools of tomorrow.

We now have the tools and practices to build this—it is time to do it.



Humans Fall in Love with Solutions—AI Can Help Fall in Love with Problems

*Why augmenting **problem exploration** with artificial intelligence may be the biggest yet underused lever for innovation*
One of the low-hanging fruits of using artificial intelligence to transform how we work is harnessing its power to help people solve problems faster, but most importantly, more creatively.

One of the main epiphanies of the last years has been how artificial intelligence can help critique our ideas, which is strengthening the idea flow at the end of its funnel - area where many people are often lacking either because of their skills or because they tend to avoid critiquing others work too directly.
Here, I want to go to the *other end* of the idea flow, the upstream part of problem-solving and creativity that, as we will see, is a key determinant of the quality of whatever happens downstream. I expand on an earlier **article** about how AI can help us discover which problems to solve. In this essay, we discuss what comes next, leveraging both current scientific understanding and practitioner experiences.
Innovation efforts often jump straight to brainstorming fixes, seduced by the "dopamine hit" of a clever solution. Innovation facilitators, for instance, know the struggle with keeping working teams focused on problem-exploration exercises, so that they do not end up paying lip service to it before moving on to the "real work". Yet theory and evidence remind us that the quality of the solution space is bounded by the quality of the *problem* space we first explore. Artificial intelligence now offers a practical, high-return way to strengthen that front-end work: accelerating, broadening, and systematizing problem exploration while keeping humans firmly in charge of purpose and judgment.

What the Research Already Tells Us

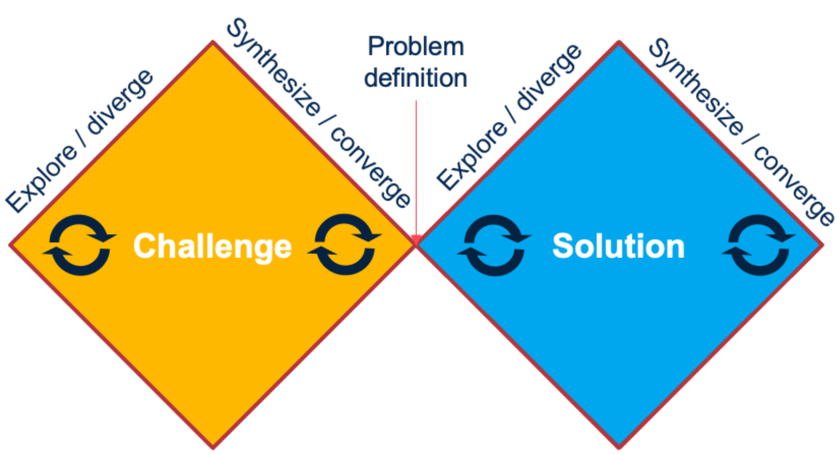
There is a reasonably extensive corpus of research on this. (To be frank, I would've expected more, but research is seemingly skewed the same way practitioners are—we focus more on solutions than on problems.)
Over the past 10–15 years, scholarly and managerial literature on innovation management has converged on the critical importance of problem-space exploration—the thorough investigation, framing, and (re)formulation of the problem itself—before moving into solution generation. Research on design thinking, creative problem solving, and strategic problem formulation demonstrates that teams and organizations that invest in clarifying and reframing the problem systematically produce more original and higher-impact ideas.

1. Time Spent Framing Correlates with Creativity
2. Multiple Problem Frames → Diverse Ideas
3. Deep User Insight Spurs Novelty
4. Reframing Techniques Lift Solution Quality

5. Reflection & Debate Outperform Rush-to-Closure
6. Problem/Solution Co-Evolution Drives Breakthroughs

Across design thinking (Liedtka, 2015; Micheli et al., 2018), creative problem-solving (Abdulla et al., 2020), and strategic management (Nickerson et al., 2012), the message is consistent: a well-defined problem is half, or at least a big part of, the innovation.

Now let's break down the issue to identify places where we can solve it. A common framework for disciplined and thorough ideation, Design Thinking's (British Design Council) Double Diamond highlights two macro phases: Diamond 1 – Problem Exploration and Definition; Diamond 2 – Solution Generation and Delivery.



Our focus is on the first diamond, where early framing determines everything that follows. Anecdotal evidence from AI-assisted ideation projects conducted through the last few years suggests material gains in speed *and* depth of insight when humans partner with AI during this phase.
A few assumptions grounded on practice can guide us here.

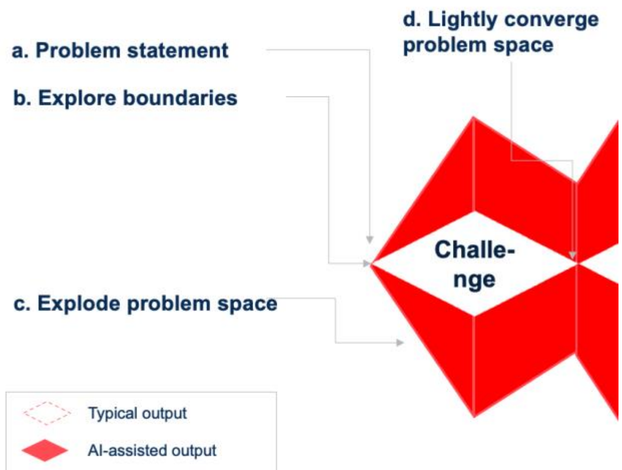
Assumption	Evidence Base
High-quality problem definitions strongly predict high-quality solutions.	Design thinking & CPS studies; practitioner consensus
Clarifying the *"Why"* and desired *"What"* of an initiative remains a human, strategic act.	Strategic management & ethics literature
Most teams under-invest in understanding the real "Why/What."	Surveys of innovation leaders; field observations

The central hypothesis of this work, supported by early evidence collected using artificial intelligence-assisted ideation technologies and practices, shows that AI, used as a cognitive partner in Diamond 1, enables a more comprehensive and insightful definition of the problem space than unaided human work—ultimately yielding solutions that are both more novel and more useful.

AI, used as a cognitive partner in Diamond 1, enables a more comprehensive and insightful definition of the problem space than unaided human work—ultimately yielding solutions that

are both more novel and more useful.

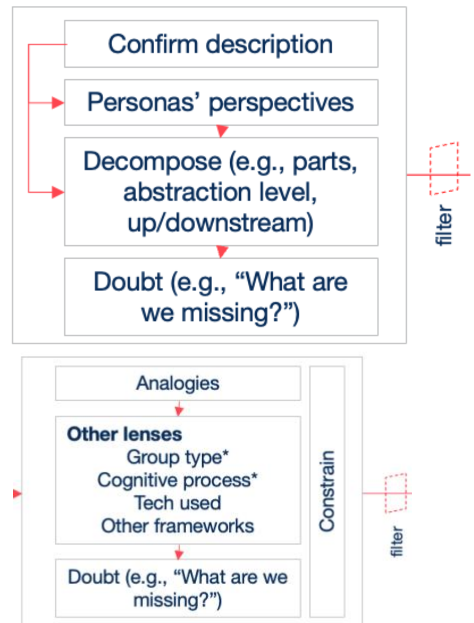
Humans still supply strategic intent and critical evaluation; the machine delivers rapid, wide-angle exploration that would be prohibitively slow or narrow if done manually.



Why the AI-Human Partnership Works

Both humans and machines are bound by our capabilities (knowledge, logic) and incentives (hormones in the human brain, tokens in the machines). Unsurprisingly, human limitations reflect themselves in organizational barriers.

AI workflows, such as those in AI chats, are aligned to follow our instinct to move fast to solutions. They also risk so-called "institutional knowledge replication," that is, staying well within the "known-knowns" instead of venturing further. However, they can be reconfigured by asking for more reflection time (and tokens) on the problem. The schemas below, part of an overall process illustrated in a previous **article**, show exercises that can help with that.



Consider the examples in the table below, where artificial intelligence can help delve into the problem and take different perspectives. Similar opportunities can be unlocked in science and R&D, among other fields.

Organizational Barriers—and How AI Helps Overcome Them

Barrier	Typical Impact	AI's Mitigating Role
Cognitive tunnel vision	Narrow frames	Generates contrasting frames fast
Representational gaps	Misaligned teams	Creates shared, data-rich visualizations
Politics & power plays	Distorted priorities	Anchors debates in objective evidence
Culture of haste	Premature solutions	Automates groundwork, freeing time for reflection

Once again, obtaining the best results requires synergy between artificial intelligence and human capabilities. People act as principled "system 2" (in Daniel Kahneman's terms) to the faster "system 1" thinking of the machines. The machines, especially if well configured and using the latest

Opportunity Area	Strategic Question Framed with AI Assistance	How the AI-Augmented "Diamond 1" Plays Out
Tariff-Exposed Order Management (Consumer Products)	*How do shifting import tariffs really affect the day-to-day experience and decision quality of our order managers?*	• LLM scans historical order logs, tariff schedules and policy news to surface choke points. • Clustering algorithms segment order-manager roles (rush vs. forecast planners). • Persona-focused prompts generate "day-in-the-life" narratives and edge cases. • Analogous contexts (e.g., pharma navigating fast-evolving covid-19 regulations) surfaced for reframing.
Decarbonizing Last-Mile Logistics (Urban Delivery)	*Which hidden frictions most impede our shift to zero-emission delivery in dense cities?*	• AI mines telematics, driver notes and city-permit data for emission hotspots. • Graph algorithms map stakeholder network (municipalities, landlords, utilities). • LLM suggests analogies from scooter-sharing rollouts and port electrification.
Clinical Trial Recruitment Equity (Biopharma)	*Why are under-represented populations disengaging before consent in Phase II trials?*	• Sentiment analysis on prior trial inquiries highlights trust breakpoints. • LLM creates culturally specific patient personas and journey maps. • Comparator scan reveals lessons from fintech KYC onboarding.

reasoning models, can also prevent us from often falling into our own System 1 thinking.
Artificial intelligence, if designed around a user experience that supports, not substitutes, the human, can help here, for instance, through:

- **Bias Buffering** – AI can surface alternative framings that counter premature convergence.
- **Perspective Multiplication** – Large models digest cross-domain data, exposing angles a single team may miss.
- **Structured Decomposition** – Algorithms break complex challenges into tractable sub-problems.
- **Multidisciplinary Scanning** – AI links insights across industries, functions and sciences.
- **Novelty Detection** – Pattern-spotting uncovers surprising anomalies worth reframing around.
- **Knowledge Augmentation** – Synthesized evidence bases raise the floor for non-experts.

Conclusion: Better Questions, Better Innovation

A decade of empirical work underscores a simple truth: the creative ceiling of any innovation effort is set early, when we decide *what problem* to solve. Humans tend to move to solutions too fast. Artificial intelligence, if instructed appropriately, doesn't suffer from the same bias or have the same dopamine hit. On the contrary, it could be incentivized to spend time understanding problems well.

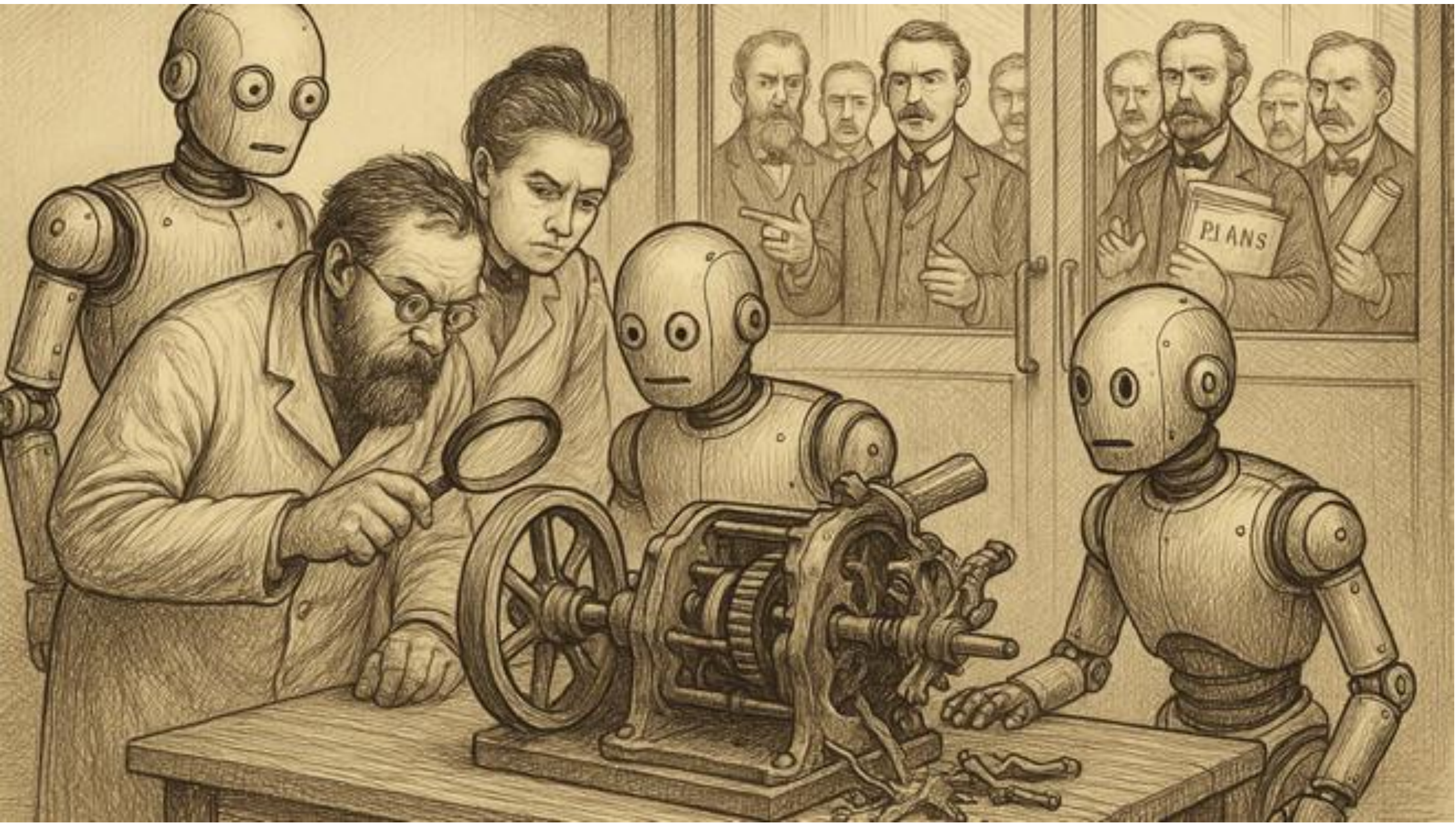
As a result, if used well, with competent humans firmly in the loop, AI now gives organizations a scalable means to deepen that decision. By pairing human strategic judgment with machine-driven exploration, teams can, among others:

- Surface hidden angles of attack, for instance, variables and stakeholder needs in hours, not weeks.
- Counteract cognitive biases that silently narrow the search space.
- Build a shared, evidence-based understanding that accelerates alignment.

The result is a richer portfolio of solution avenues and, ultimately, more original and valuable solutions. Companies that cultivate disciplined, AI-augmented problem framing are not just "doing design thinking faster;" they are upgrading the very substrate of innovation, ensuring they invest in solving the right problems before investing in solving them well.

Abridged References

- Abdulla, A., et al. (2020). *Problem finding and creativity: A meta-analytic review.* **Psychology of Aesthetics, Creativity, and the Arts, 14**(1), 3–14.
- Baer, M., Dirks, K. T., & Nickerson, J. A. (2013). *Microfoundations of strategic problem formulation.* **Strategic Management Journal, 34**(2), 197–214.
- Carlgren, L., Rauth, I., & Elmquist, M. (2016). *Framing design thinking: The concept in idea and enactment.* **Creativity and Innovation Management, 25**(1), 38–57.
- Dorst, K. (2011). *The core of "design thinking" and its application.* **Design Studies, 32**(6), 521–532.
- Johanssen-Sköldberg, U., Woodilla, J., & Çetinkaya, M. (2013). *Design thinking: Past, present and possible futures.* **Creativity and Innovation Management, 22**(2), 121–146.
- Liedtka, J. (2015). *Linking design thinking with innovation outcomes through cognitive bias reduction.* **Journal of Product Innovation Management, 32**(6), 925–938.
- McKilligan, S. & Creeger, S. (2018). *Strategies to redefine the problem exploration space for design innovation.* Proceedings of **E&PDE 2018**.
- Micheli, P., Wilner, S., Bhatti, S., Mura, M., & Beverland, M. (2018). *Doing design thinking: Conceptual review, synthesis, and research agenda.* **Journal of Product Innovation Management, 35**(5), 740–758.
- Nickerson, J. A., Yen, C. J., & Mahoney, J. T. (2012). *Exploring the problem-finding and problem-solving approach for designing organizations.* **Academy of Management Perspectives, 26**(1), 52–72.
- Seidel, V. & Fixson, S. (2013). *Adopting design thinking in novice multidisciplinary teams.* **Journal of Product Innovation Management, 30**(S1), 19–33.
- Studer, J., Daly, S., McKilligan, S., & Seifert, C. (2018). *Evidence of problem exploration in creative designs.* **AI EDAM, 32**(4), 429–444.


PLANS

Harden Your Ideas with AI

Big and small ideas are rarely perfect when they first come out. The best idea creators and problem-solvers have solid processes for fast and cost-effective improvement. However, most people aren't in that category. In particular, most people aren't great at looking for feedback, receiving it when given to them, or giving it to others. The feedback and iteration phases are some of the trickiest steps in the innovation process.

This need and challenge apply to any idea—from the simple "find me an existing tool for well-understood problem X" to anti-disciplinary exploration for poorly defined systemic challenges and anything in between. Arguably, most of us should systematically get feedback on more of what we do before we do it.

Enter AI. If used well, AI allows us to generate many more ideas, alone or in small or large groups. The opportunity is enormous. But so is the risk of generating a mountain of half-baked platitudes that humans find hard to sift through.

In some cases, AI is already pretty good at filtering ideas based on novelty and sometimes even desirability, feasibility, and viability. It becomes even better so when guided deliberately by humans. These are promising directions, but they're unlikely to solve the problem in isolation.

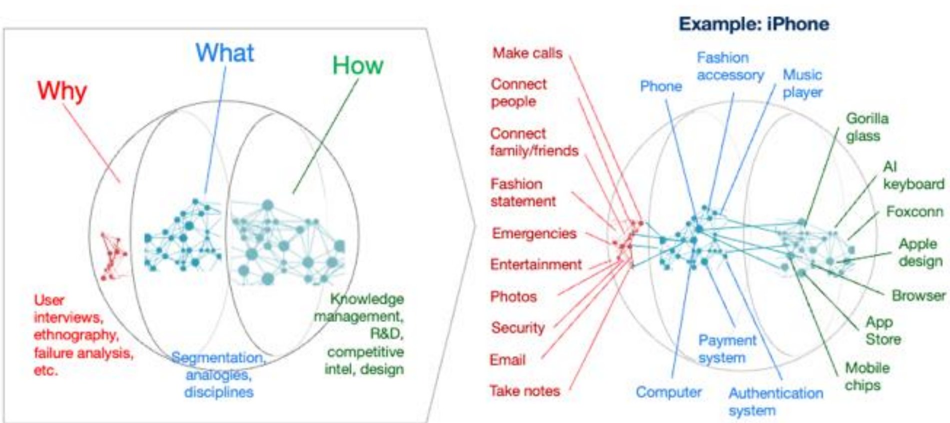
Beyond AI as a solution generator

Back to the importance of feedback - What about enlisting AI as a partner in critiquing ideas so humans (and possibly other machines) can iterate? We intuit that AI can ask questions, not just provide answers. What else can it do to spar with us and improve our thinking?

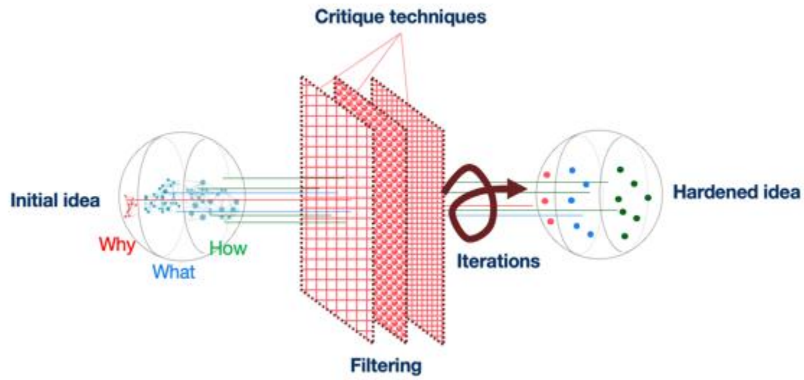
For one, current GenAI machines are already pretty good at taking different perspectives (for example, personas), looking at problems from the standpoint of various disciplines - if asked to do so - and providing enough context when needed. However, they may struggle with symbolic reasoning, abstraction, and conceptualizations that require a deeper understanding of how the world works.

This is where humans are helpful. Both individually (you and I critiquing someone's output) and, most importantly, in our established-knowledge avatar. That is, guiding AI to sift through and use the myriad artifacts we have built over hundreds of years of crystallizing our reasoning into theories, frameworks, and practices. We can point AI at those and use them as lenses, combining AI's brute force with the *symbolic* reasoning hundreds of generations of competent humans have distilled. In a way, drawing from Kahneman's *Thinking Fast and Slow*, humans (and our collectively-produced artifacts) add System 2 thinking to machines' System 1.

Think about it: combining GenAI's strong language ability with language and human models of the world, we can attack ideas at multiple levels: their "why" (the reasons why something should be done, typically addressed through some form of research), the "what" (the categories of possible solutions, beyond the obvious ones, such as analogies), and the "how" - which is typically what most people and machines individually would pay the most attention, drawing from existing "how-to" sources. See the example below.



Such a richer representation of the problem/solution space is often the key to identifying novel solutions. An Idea Hardener can help with that: it can combine human frameworks with AI's brute force and boundary-free knowledge by making better-described ideas undergo multiple layers of inquiry. That architecture is based on collective intelligence principles augmented by AI - not AI throwing up incomplete or naive solutions.



An example of the practical implementation of an "**Idea Hardener powered by ACI**" (Augmented Collective Intelligence) is available on the OpenAI GPT Store. It is a prototype usable in various contexts and can also be configured or customized for specific organizational environments. The tool consists of a principled, systematic process of exploring and applying lenses based on nearly one hundred well-understood methods. Special care was taken to collect and harness myriads of critique methods from multiple disciplines, in an exercise of harvesting collective intelligence. Some examples: challenge and opposition techniques (e.g., red teaming, motte-and-bailey test), failure and risk prediction (e.g., pre-mortem, Failure Mode and Effects Analysis (FMEA)), root cause and decomposition techniques (e.g., 5 whys, first principles thinking); future and scenario exploration (e.g., Delphi method, backcasting); perspective shifts and reframing (e.g., inversion thinking, defamiliarization); and many others, including ones that probe into ethics and fairness.

In the qualitative observations from the tests done so far, conservatively, the Idea Hardener allowed users to generate as much relevant critique in 15 minutes (the typical run) as a team of 3-4 people could produce in 1-2 hours. It also provides perspectives from more fields than typical humans, including professionally trained innovation facilitators.

What's more, the concept is extensible. Individual organizations can build their own with personalized processes and data sets. Additional use cases can be performed, such as preparing for an interview or an exam. Multi-modal (video, audio) capabilities are becoming ubiquitous and could be incorporated here.

A friendly but honest critique can buttress many things.

The upshot for people and organizations

This or similar tools *aren't intended to substitute humans* for giving feedback—keeping good human critics in the loop is essential, as their intuitions may be very insightful and more deeply consider elements like empathy and fairness. In general, trained, competent humans add diversity and more than a note of originality to AI's output. There is a risk that people would overly rely on AI Idea Hardeners and forgo doing their hard work to critique things, potentially leaving the door open to bias.

However, such tools can *complement* people by offering a scale, scope, and speed that typical ideas and problem-solving processes can't afford.

In other words, they could yield at least three significant advantages if used correctly (e.g., as a complement, in parallel, and downstream from *human* feedback).

- **We could harvest more feedback.** First, they increase the absolute **amount** of critique performed, as people (including individuals working independently) who wouldn't do enough of it for lack of time or capability now have an always-on specialized assistant to help them. Tire-kicking could happen more pervasively across organizations.
- **We could make people better at it.** Second, they **strengthen the ability of** problem-solving and innovation **teams** to avoid blind spots and groupthink and help them learn these techniques by working with the tool.

- **We could genuinely "fail faster."** Third, they increase the **speed** at which teams can weed out bad ideas, reducing the overall cost of failure in innovation pipelines (including those that use AI to transform work).

It is time to embrace AI-powered idea hardeners. Do you have yours? 🧠

Relevance is (much of) what we need from AI

Most of us can solve complex problems because we get *the right external input at the right moment*, often over long periods. That typically comes from our colleagues or things we see and read, and it can be amplified by AI—some of it today, a lot more, very soon. In the long term, the picture of human augmentation is even more intriguing. Let's look at what can be done.

The future of agents and copilots is one where "**whispering machines**" see what we do from our screen and our environment (both of which could be helped by new multimodal AI features), and with that context, they inform their suggestions.

Imagine them proactively whispering new *relevant* ideas, discussed in articles or blogs, as input for what you are actively researching or learning. Relevant ideas are not just ideas and knowledge in spaces we care about - they are useful to what we are *actively trying to do and think about* (right now or on a longer time horizon). They're not duplicative of things we already know or that we have already surfaced. That ability, technically largely feasible today and poised to become even more so as AI context windows increase and AI assistants learn from what we individually do, will help humans filter through the immense amount of noise. It is the digital, massively scalable equivalent of layers of perceptrons before our human eyes. A sort of bionic augmentation that we can use right now. We try to do that job today by manually finding sources and following them by reading, listening, and watching as much as we can, either individually or with the help of our teams and organizations. Some tools exist, such as Feedly or Curata. But we are increasingly unequal to the task—we miss out on useful knowledge and inspiration, and that infinite treadmill tires us.

This feels like (in Christensen's words) a "job to be done" that could truly benefit from machines that both sense and understand/remember our context and can effectively sift through large amounts of multimodal knowledge. And that doesn't mean "enshittification," i.e., flattening of the output towards mediocre, unoriginal ideas: we could instruct the machine to explore analogies and adjacent fields and bring back ideas from those fields. One upshot, among others, is to mitigate the effect of misinformation: we could enlist machines to filter out egregiously wrong things by applying logical filters and subsequently would support us in deciding what to believe by asking us questions to push our critical thinking.

Communities and their (in MIT's Malone words) *superminds* can also play a large, exciting role in collectively providing context to machines instead of doing that as individuals. From an augmented collective intelligence standpoint, the time is ripe for AI to support this specific pillar of a supermind's architecture: **information feeds**.

Finally, we can expect *knowledge graphs* to play a big role in connecting people's thinking and disparate pieces of content in a more principled and symbolically-minded (conceptual, as opposed to just semantic) way.
How could that play out in 2030? Here's a short story based on all of this.

Smartstreams 2030

I open this morning's smartstream. I am interested in regenerative agriculture, and the personalized AI curator summarizes what people in comparable climates have done in the last weeks. So much time is saved; this is magical.
Smartstreams provide curated content and identify relevant people to follow and engage with. They are an offspring of the old social media and have made up for their forebears' spotty track record. I can finally fully configure my information diet and be current on the latest.
It is now easier to engage with people in fields I care about because noise (trolls, uninformed opinions) is filtered out. However, to prevent insularity, it is now a legal requirement for service providers to "mix things up": the algorithms must inject some dissonant opinions - if they're civil. It is not too hard to do that now that we have understood how to combine natural-language search with the network signature of "dissonant voices" for a given topic, including the most arcane ones.
We now pay for quality content instead of assuming that good content should be free. Governments and private catalyst capital have finally stepped in to subsidize a minimum of quality content for people who can't pay, especially for sensitive themes that lend themselves to dis/misinformation (it just took a populist scare and a few years of political mayhem, but hey, we are here now).
Smartstreams curate possible answers but also transmit questions. Within companies, Smartstreams convey the questions asked by other colleagues: they are automatically tagged and routed to the right people instead of requiring insider knowledge of the firm. Some communities have enabled that feature across all members, irrespective of their organization.
By law, smartstreams must notify users when they seem to be inducing signs of addiction and dependency. My wearables combine feeds from my eyes, my brain, heart, blood, adrenaline, and other signals and tell me, at first softly and then firmly that I am overdoing a couple of rabbit trails this morning.
Smartstreams are complemented by a new generation of search engines—the recent bit of progress I am most thankful for. Some of the technology had existed for a few years but wasn't seamlessly combined and wasn't monetized

effectively, resulting in unhelpful bias towards what advertisers would find useful. No more—or at least for the premium search engines that many were waiting for.

I formulate my question, and the natural-language AI coupled with a deep knowledge graph helps me rephrase it to avoid blind spots the way an expert librarian would—a librarian who knows the domain I am searching for every major domain. Then, the engine breaks down my question's semantic and symbolic space. Concepts are mapped in a two-dimensional space, so I see insightful adjacencies.

Caring about the question is as important as caring for the answer. AI helps there.

Then, some more magic is in store. The engine summarizes the results and gives me highlights and synopses of the main sources for the answer, e.g., scientific papers and reputable articles—including information about unresolved disputes. It displays results in a two-dimensional graph, so I can look at the periphery, where inspirational, edgy nuggets lie.

The engine over-indexes credible information based on what is said (triangulating it across sources) and who says it, as well as what the web knows about that person's (and their network's) credentials.

If I override that setting and look for more "unusual-suspect" views, the engine tags the results with a reliability—and possible harm—score.

Next, it displays the underlying network of people and organizations and asks me if it should post my question to them. Last, I add some of it to one of my smartstreams, to track the developments.

Some of this was possible within enterprise networks earlier but not across the entire web. Much paywalled research is also now searchable, thanks to a combination of free preprints, philanthropic capital, and public contributions to the knowledge commons. Language barriers are now irrelevant, as everything is translated.

The upshot? The cycle of invention-to-scale is shorter. With some of these ideas, we can feed our Idea Colliders in real-time, and hyperspecialized knowledge communities are being built on the search APIs. For example, an array of climate-transition superminds—the kind of stuff we badly need in 2030.

Better Ideas by Taking Turns with AI

As AI becomes more integral to our work and lives, learning how to collaborate effectively with these tools is increasingly important. Yet, there's still a fair amount of confusion. Some people, sometimes unconsciously, treat AI as a mere content-generating machine, expecting it to drive the reasoning while they sit back and watch. This passive approach can undermine creativity and critical thinking.

Interestingly, this phenomenon isn't entirely new. We've all been in meetings or workshops where a few confident individuals dominate the conversation—sometimes management consultants, sometimes enthusiastic tech folks—leaving everyone else hesitant to contribute. When certain voices overshadow the group, creativity can suffer. In this short essay, I will focus on a simple, yet not trivial, behavior change.

A proven method to counter this dominance in brainstorming sessions or design-thinking workshops is "brainwriting." Before anyone speaks, participants take a moment to jot down their ideas. This levels the playing field, encouraging contributions from those who might be more reserved or easily distracted by louder voices. Brainwriting ensures everyone's input is captured and no single perspective dominates too soon.

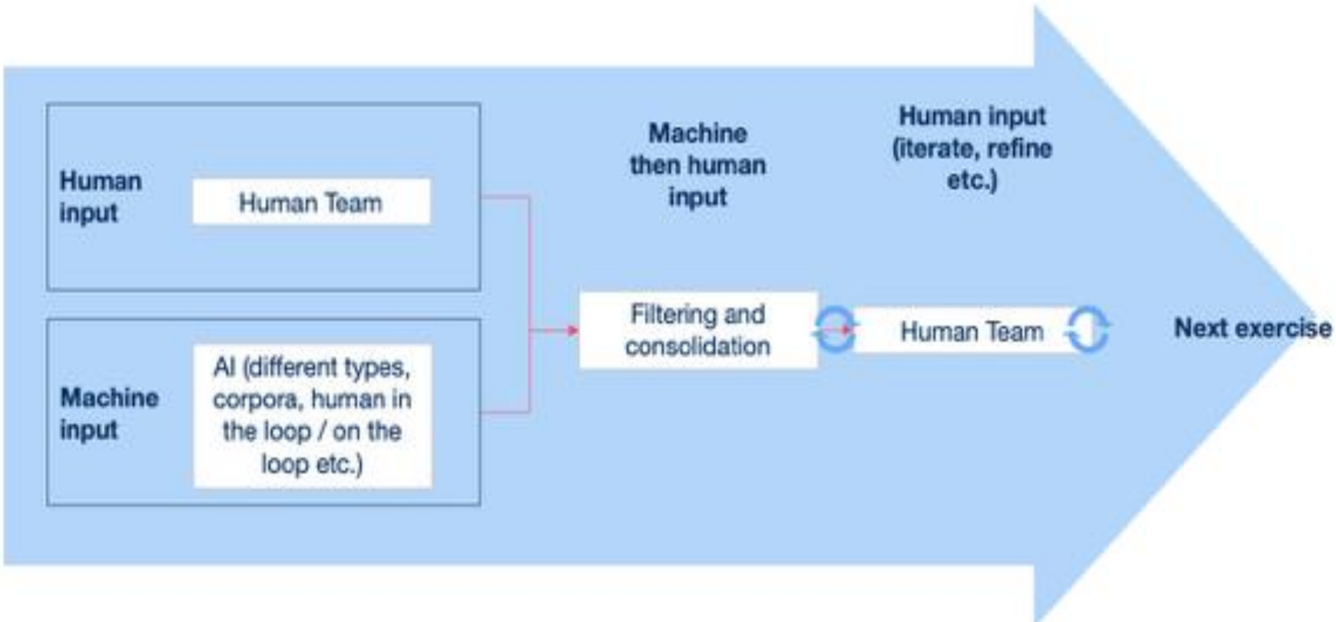
Working alongside AI can feel like adding another very smart, and somewhat imposing, member to the team.

As in a typical group setting, managing how (and when) ideas are introduced and built upon is essential. One effective strategy is to document your own thoughts first—maybe even have the AI ask clarifying questions—before letting the AI propose solutions. This prevents overreliance on AI-generated content and keeps humans in the driver's seat of creativity.

Imagine a problem-solving workflow like this:

1. **Human Input:** You begin by outlining your ideas, questions, or goals. This can be done individually, but it is better as a group, with each person writing first.
2. **AI Input:** Then (or in parallel), you ask the AI (or multiple AI systems with different knowledge bases) for its perspective.
3. **Filtering & Clustering:** The AI can help filter and cluster ideas—both yours and its own—into coherent themes.
4. **Human Review, Iteration & Synthesis:** You and your human colleagues iterate, refine, and build on these combined ideas, integrating your professional judgment and creativity, and potentially using AI to identify improvements or provide further inspiration.

The chart below represents this flow schematically.

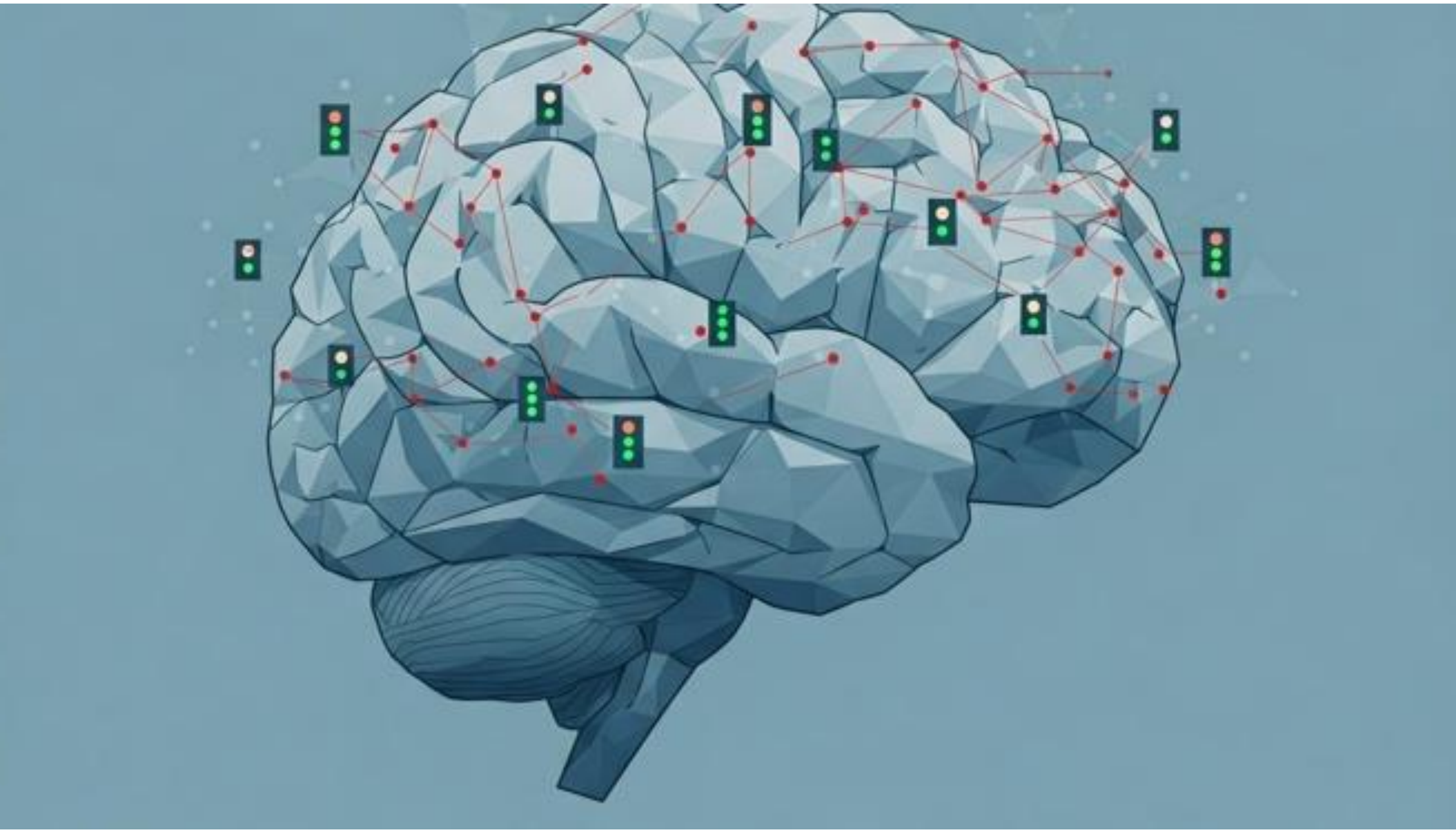


This process is rarely strictly linear; you might loop back to the AI after additional insights emerge. Also, in some scenarios, it might make sense for the AI to offer an initial fact base—like "pre-reads" that help everyone on the team start with the same information. This may be particularly helpful when discovering poorly understood problems.

Either way, maintaining a sense of control and ownership over the flow is crucial for better outcomes, preserving morale and motivation, and keeping crucial parts of the human brain activated.

Ultimately, collaborating with AI should feel like working with any other skilled teammate: you give input, receive input, and find synergy by taking turns. We must develop this habit as a behavior change—actively shaping the conversation and not just reacting to AI suggestions.

If done right, augmenting our intelligence with AI can spark innovative solutions that neither humans nor AI can reach alone. If done wrong, we might end up with lots of boring ideas, and our own ability to think creatively might suffer. This is not a philosophical choice; it is about applying the right management practices to the new world of work.



“AI Psychedelics” For Radical Innovation?

Radical innovation often stems from bending reality, a practice deeply ingrained in human history. Psychedelic drugs, embraced by some, including artists and visionaries like Steve Jobs, and currently intensely researched in the medical field, have long served as one avenue for this exploration. Can AI, especially its generative type, perform some of the same functions on our *collective* brain, consisting of the cognitive interplay between many of us?

Where radically new ideas come from

First - how do radical new ideas get generated and refined? There is no one path, but let’s start by pinpointing some core tenets of innovation and its creative process:

- An **unusually intentional focus on problems** to empathize (“fall in love”) with them before any problem-solving attempt
- Discovering **connections** between traditionally separate ideas to uncover unconventional solutions
- Using **constraints** as filters and redirection
- Soliciting thorough, possibly uncomfortable **feedback**
- **Iterating** relentlessly.

These principles often challenge individuals, especially when venturing beyond their expertise, or trying to scale efforts. They may also be in the strike zone of psychedelics’ effects - both the physical and the digital version of them.

Human brains on psychedelics vs. collective intelligence "on AI"

When effective, psychedelics affect the human brain in many ways. They can significantly alter both functional connectivity (how the brain’s connectivity functions) and structural connectivity (how the brain’s connectivity is physically structured). Notable effects include disrupting the default mode network (DMN), which is linked to self-referential thoughts and mental chatter and leads to altered self-awareness and unique perceptions. Psychedelics also increase crosstalk between brain regions, fostering unusual associations.

Psychedelics facilitate deep attention to otherwise overshadowed details (typically drowned by ongoing, routine brain activity) and encourage idea recombination while suspending judgment—similar to design thinking principles.

Now, consider what generative AI can achieve in a human-machine creative (work)flow. It can potentially transcend our insular, self-referential individual and group thinking - the collective equivalent of DMN chatter, enabling the exploration of concepts from diverse fields. Teams and processes, for instance during workshops, have traditionally aimed at such exploration. Networks and ecosystems of people cross-pollinate and help new ideas germinate. Now, increasingly smart AI can augment their collective intelligence.

In some ways, weaving AI into this process mirrors how controlled psychedelics interventions require pre-session preparation and post-session integration (guided reflection) with trained coaches for a transformative experience. Steve Jobs and artists who used psychedelics did so as part of a creative process that involved creatives, domain and technology experts, creative bursts involving AI demand preparation and integration. *Human involvement remains integral to the ideation process, grounding and directing AI's exploration*.

How to get there

Here are some ideas to help design your collective AI-human creative workflow:

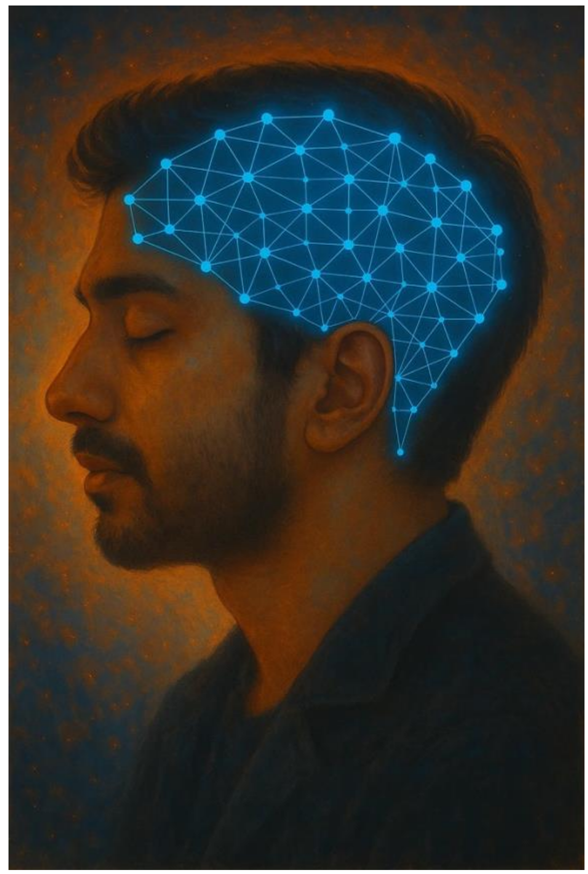
- Push AI to thoroughly **explore problems**, with your team's input, instead of assuming you and the machine understand it upfront
- Utilize the crystalline of **frameworks** built by unusually insightful humans such as management theorists, creatives, or scientists
- Ask your AI and human teams to identify **analogies** that shift the frame of reference to foster unique perspectives
- **Inject ideas** from other fields, and ask AI to recombine them in unusual ways, possibly with your team's help
- Force the machines to **take perspectives**, possibly based on specific personas, and generate dialogues that expose the innards of the problem in new ways
- Verbally encourage AI's **edginess** and unconventional solutions
- Ask AI to **iterate** with those different perspectives in mind, and filter or cluster the results to avoid overwhelming human ability to judge and complement them.
- Possibly, and certainly, in the future, explore the use of **knowledge graphs** to guide systematic recombination across fields for the exploration of both problems and solutions.

Potential side effects, positive and negative

Just possibly, this approach might also help professionals break free from self-referential, siloed (and innovation-inhibiting) thoughts, similar to the potential and currently much-researched long-term benefits of psychedelics for depression and anxiety patients.

On the flip side, there is a risk of dependency, that is human over-reliance on machines' ideas, and a corresponding reduction of independent, unassisted creativity. (As a side note, psychedelics don't seem to generate physical dependency, unlike many other drugs, but the consumption patterns exhibited by creatives who might have found them an easier way to create might be more intense). The design of UI/UX can help address some of it.

The Bigger Picture: Impact on Society



Will AI Sharpen or Dull Our Minds?

Originally published in February 2024 on Exponential View by Azeem Azhar

AI is percolating into our economy and society, and it surrounds humans in a way it never did before. It has become ambient. Is that good for our intelligence?

Some highlight the risks. The *FT*'s Tim Harford recently asked "will we be ready" to assist the AI when it needs our judgement to make a decision? Or will we get stuck in the "paradox of automation," where humans lose the ability to intervene when AI systems need us to? Some scenarios are benign, but many others are existential: like pilots over-relying on automated flight systems only to crash the plane when the computer goes dark (see, for example, the tragic case of Air France 447).

In this first commentary, I will break the question down into two:

1. **What is the risk for the individual?** (a) The risk of becoming less attentive, less critical, less creative, less proactive? (b) The risk of not developing some foundational skills anymore? Is AI going to deprive us of some learning by doing?
2. **What is the risk for our collective intelligence?** This is not about the average or total of our individual intelligences but rather the *emergent* intelligence capabilities of the *structures* made of networks of people and assets (including machines) that behave collectively in ways that show intelligence above and beyond that of the individual components. Is that going to improve, or worsen?

Individual risks, and rewards

It stands to reason that some of the downside risk is real. But is it inevitable? And what is the upside?

The net effect of technology introduction has been in the past economically (and typically socially) positive in the long run. For sure, there can be huge volatility, and indeed dislocation, that sometimes last a long time. In exponential scenarios, with potential systemic instability, the past may not automatically be a good predictor of the future.

The research doesn't seem to be fully settled, but there is some, and we can frame the problem based on a few examples:

- The invention of **agriculture** didn't make individual people smarter than hunter-gatherers. Some research even indicates that the size of our individual brain might have shrunk as our collective one, emerging from our societies' networks, grew. BUT: without agriculture, the world's society would likely be more primitive, and most of us wouldn't want that world today.

- The introduction of the printing **press** might have reduced most people's ability to recite books by heart, and even contributed to the disappearance of jobs such as professional storytellers. But the effect on individuals (printed materials aid cognition) and societies (knowledge management) was a net positive.

- **Taxi** drivers in London, after GPS introduction, didn't have the same quality of spatial reasoning (and even their brain structure changed). BUT - did that make them worse taxi drivers? It seems to have helped less experienced drivers become more effective.

- **Typewriting** was bad for handwriting (and handwriting is likely related to some level of creativity), BUT that was more than offset by other gains. By some accounts, typewriting saved forty minutes out of an hour, compared with the pen. Automated orthography corrections are increasingly making us unable to thoroughly spell-check things alone, BUT that allows us to write more.

- A study on the use of robots in helping baseball **umpires** shows that the combined human-machine duo improves performance over humans alone, especially for lower-skilled humans. Humans who after using robots don't receive assistance anymore seem to not be able to get back to their original skill levels. BUT: The introduction of robots also makes the game less acrimonious, with fewer disputes and expulsions. And a recent study on the use of computer vision in tennis showed that human umpires show better judgement when technology is deployed alongside them.

- Even where humans lost the battle, like in playing **Go**, evidence shows that machines' superiority pushed up the quality of the average professional Go player. After all, a game is supposed to be making us better - and in this case, AI competition did.

What about not developing some foundational skills?

Is AI depriving us of learning by doing?

How do we create stepping stones in some professions when machines do a lot of the entry-level work?

Consider modern finance, legal, and consulting professionals who haven't developed, respectively, the algebra, writing, or handwritten storytelling skills of their predecessors. Does that make them less intelligent, or did that rather force them to develop skills that built off those machines, and spend more time on other tasks, such as interfacing with their stakeholders?

One transferable example comes from an unexpected place. About 10 years ago, there was a big concern in the Finance/Accounting community about the fact that the Finance Operations jobs were increasingly centralised in low-cost locations or outsourced, which means that future Chief Financial Officers (CFOs) wouldn't grow up professionally by doing low-level work and then moving up. Ten years later, we don't talk about that so much. For sure, some of the old skills, like the ability to spot mistakes in accounting systems, might have dwindled. Exception management, including its data mining and analytics component, instead of the daily running of operations, is where finance executives get trained for the top job. And indeed, they now learn how to have separate organisations run industrialised operations - as if they had their own supply chain. The new aspirant CFOs also have plenty of room for other capabilities that they can do more of: focusing on the crafting and the execution of strategy, sustainability, and partnering more closely with their peers and their organisations in running the business — as well as, of course, learning how to use advanced analytics and AI. Those who have embraced the change now thrive.

Humans have historically adapted to the introduction of new technological tools by developing new capabilities that complement those tools and push productivity - writ large - higher. At least, they did it so far, and in the long run.

What about the collective brain?

The collective intelligence side of the story shouldn't be conflated with the previous one. From the printing press to the telephone, from email to the internet, and from mobile phones to Google, the introduction of collective-intelligence-enhancing architecture has historically enabled an explosion of collaboration and substantially reduced the time to access new knowledge. As a result, our knowledge graphs, with both content and people as nodes, new relationships have changed and their edges are now able to connect more ideas than ever. At parity of individual intelligence, on balance, that has made us - and certainly could make us - collectively smarter.

At the same time, algorithmic curation optimised on human tendencies has possibly deteriorated our ability to function cohesively as a society (see social media discourse polarisation, and at least partially related social polarisation - especially in the US), and likely impaired the resulting decision-making (political governance, or lack thereof, come to mind). The interplay between our godlike technology, Palaeolithic brains and mediaeval institutions[1] might very well not lead to a net-higher collective intelligence today, at least in the short run. There is a real risk of dulling our supermind, right here. We will know even more after the many elections of this year.

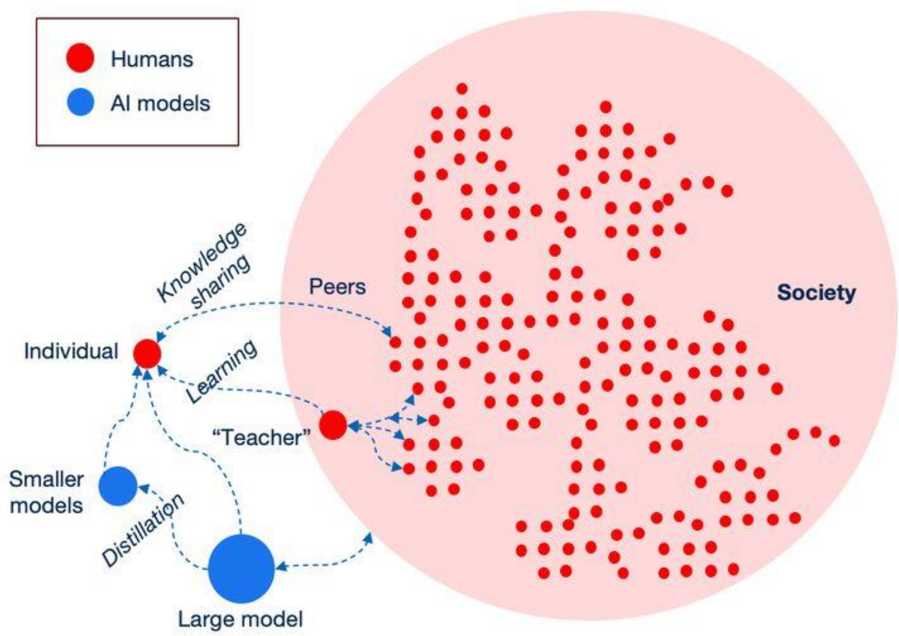
Enter generative AI, and its alluring confidence, its ability to spin gratifying new artefacts in seconds, effortlessly. There is a real risk that many, too often, would get hypnotized, lower our guard and not exercise quality control. Some evidence points to humans "falling asleep at the wheel". When the LLM made mistakes, BCG consultants with access to the tool were 19 percentage points more likely to produce incorrect solutions. And the range of ideas generative AI produces out of the box is not as good as what humans, collectively, would produce. Microsoft recently published a good literature review of the dangers of overreliance on AI.

It is ours to shape

So the risks are real, but they don't seem unavoidable. In the next essay, I will explore the solutions available to us today, and some frameworks to keep developing them as capabilities - human and technological - change.

Are we small models?

Something is afoot, and we may be missing the big picture, especially as large models start using distillation methods to train cost-efficient small models, following DeepSeek's lead. For thousands of years, humans have learned from peers and teachers (of all sorts), effectively leveraging them as curators, filters, and lenses for the broader world's knowledge. New knowledge management methods, such as the printing press and computers, multiplied that ability to learn from the world's "supermind". What is happening now is that the world is teaching AI large language (and multimodal) models, initially through scraping and annotation, that in turn, through distillation, teach language models that are smaller and more specialized, but able to perform specific tasks at high quality and very low computational cost. What strikes me is that individual humans like you and me are a sort of small model whose learning has been distilled through interactions with entities that collectively crystallize the world's knowledge. So, in a way, the world's collective intelligence is the large model that teaches us. And now it gets augmented by, and partially flows through AI tools, whether large or small models with which we interact. The increased capabilities and usage of artificial intelligence, combined with advancements in knowledge technologies such as knowledge graphs, will push this to new limits, including influencing our own learning.



I am unclear about where this will end—there is both genuinely great potential and real risks. For instance, if tomorrow's knowledge is filtered first by machines before it ever reaches us, who gets to tune the filter?

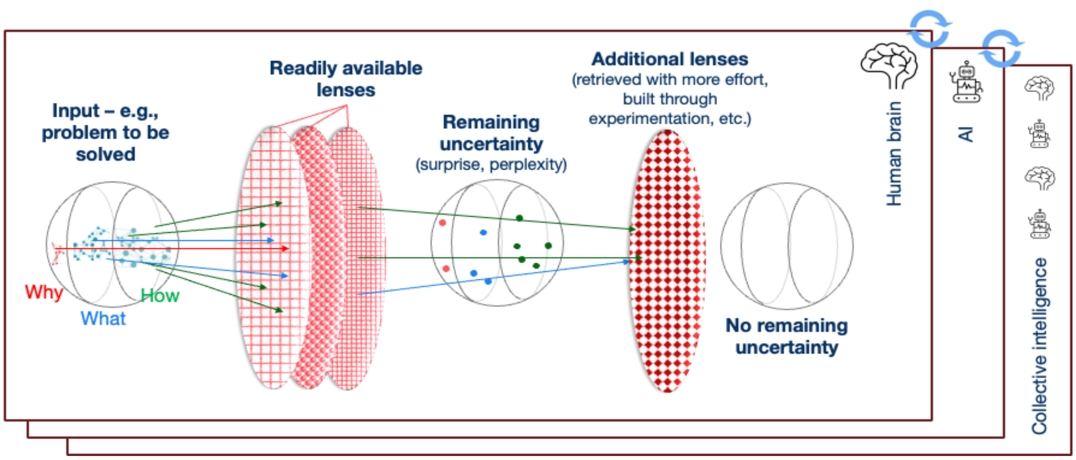
The discussion is to be continued.

Us and our machines are lenses – and that matters immensely.

Another inflection point is sneaking up on us, as many are fixated on AI's technical side. Think of the following, which doesn't fit well in a single discipline, except in the emerging field of AI-augmented collective intelligence.

Two intuitions come from neuroscience and computer science.

Neuroscience's active inference is Karl Friston's idea that brains expend additional energy to reduce the gap between their mental model and reality, thereby minimizing surprise, which is beneficial for survival. Children, for instance, do a lot of that, and that's likely why their neural infrastructure is more voluminous than adults until it starts pruning itself in early adulthood.



Computer science self-supervised learning involves machines shaving entropy off data until a pattern clicks and dimensionality (the rough equivalent of thinking "dog" instead of "fur, ears, tail, paws, etc.") reduces, which is helpful for effectiveness and efficiency).

Put the two together, and you get a single idea: any cognitive entity builds models of the world - it builds "lenses", frameworks through which the world can be understood efficiently, connecting dots instead of needing to make sense of all dots. And at runtime, it uses the most quickly available ones, simpler and energetically cheaper, dimensionality-reduced heuristics and shortcuts first – but then switches on to a "system 2" thinking if it doesn't feel that the answer is good enough, hence spending more time on the problem and potentially building new lenses on the fly.

Every framework our species has invented and stored through language and increasingly sophisticated knowledge management—Newtonian mechanics, Lean Startup, first-principles finance—acts like a pre-built lens. It compresses messy reality into "good enough" predictions with minimal cognitive wattage. When you ask an LLM to explain Porter's Five Forces or TRIZ, it's yanking that lens off the shelf for you.

Are We Lens Libraries?

Individually, we hold a quirky subset of civilization’s lenses: your coach’s feedback, your grad-school stats model. Collectively, as we lean on one another (our colleagues, friends, communities, families, and societies), we form a giant, distributed library that remains adaptive because each node continually experiments and shares.

Can Machines Augment That? Generative AI models already pattern-match; the next jump is "lens orchestration":

- Meta-selection – Choose the smallest lens that collapses uncertainty well enough.
- Lens fusion – Stack multiple lenses. Say, option pricing + behavioral econ + climate data.
- Lens evolution – Run simulated or real-world experiments, score the lens, mutate it, and redeploy. Part human, part machine experiments.

Humans Still Matter Here. Symbolic reasoning (the ability to mint an entirely new lens) is still our comparative advantage. But we’re terrible at recall and consistency; machines are fantastic there. And machines can combine things for us, at scale. That is, if we guide them to deliberately use lenses, not just single-shot next-token prediction.

So next time you work with GenAI or their agentic counterpart, work through deliberate lenses, which you can choose and blend with AI's help. AI will typically do a good job because it can "interpolate" between many examples of lens use, often across different fields.

And the bigger picture: Can we build "perpetual motion machines" where AI seeks patterns to test (including with us) and make new lenses? An AI-augmented collective intelligence, also known as a **supermind**, can lead to perpetual serendipity and connect the dots, allowing other humans and machines to leverage the results effectively and efficiently.

We Are GenAI's System 2

The world is trying to understand the potential of Generative AI, and many resources —most—are going into improving AI models. However, exploring how human-machine collaboration can enhance accuracy and insight is also helpful.

One promising direction is leveraging Daniel Kahneman's System 1 / System 2 Thinking, distinguishing more intuitive and faster thinking modes from more reflective and slower ones. While AI companies are considering this framework in their algorithm-enhancing research, I want to focus on the immediate opportunity for most organizations and users: AI-augmented *Collective* Intelligence (ACI). That means ensuring humans are in the loop as System 2 to complement machines' System 1 (and with the new OpenAI GPT o1 model, possibly "System 1.5").

Thinking fast and slow is how cognition happens

Kahneman's model, from his seminal book *Thinking, Fast and Slow,* distinguishes between two modes of cognitive processing:

System 1 Thinking:

- **Fast, automatic, and intuitive:** This mode operates almost effortlessly, drawing on instincts, emotions, and past experiences to make quick decisions. In moments of stress or danger, this also means keeping us out of trouble. Like driving your car and making quick decisions if something unusual happens.
- **Unconscious processing:** it functions beneath the surface, handling routine tasks and swift judgments without deliberate thought.

System 2 Thinking:

- **Slow, deliberate, and analytical:** This mode requires conscious effort and is used for more complex problem-solving and decision-making. Like finding a win-win solution to a complex negotiation with a supplier.
- **Logical and rational:** System 2 engages when we need to consider information, analyze data, and weigh options carefully.

These aren't separate brain systems but conceptual models that illustrate how we process information. They often work in tandem, influencing each other and sometimes operating simultaneously. While both can be prone to errors and biases, System 1 is more subject to them because of its speed.

This is also not a watertight divide. Specific System 2 processes can become more automatic with practice, resembling System 1's efficiency. Moreover, the distinction between these systems is more of a continuum than a strict divide. This said they are helpful as a principle for what we need.

Computers are faster. Humans can make them more logical and deliberate

In the context of human-AI collaboration, integrating System 1 and System 2 thinking offers a helpful framework:

AI's System 1 Input to Collective Intelligence

- **Routine Tasks and Automation:** Just as System 1 handles routine tasks automatically in humans, AI can efficiently manage repetitive tasks such as data entry, sorting, or preliminary data analysis. This automation frees humans to focus on more complex challenges.
- **Instantaneous Responses:** AI provides quick, heuristic-like responses to straightforward queries, mirroring System 1's rapid decision-making. It will likely tend to choose standard, "safe" answers. For instance, this capability is particularly valuable in customer service or real-time data monitoring.
- **High-Volume Pattern Recognition:** AI’s strength in identifying patterns in large datasets parallels the intuitive pattern recognition of System 1 thinking. For example, in market research or employee experience analyses, GenAI can identify conversation patterns across many respondents, enabling analysts to engage with that corpus more effectively. Clearly, GenAI gets many patterns wrong, which requires humans to be in the loop (more below.)

Humans' System 2 Input to Collective Intelligence

- **Complex Problem-Solving:** AI can process vast amounts of data, distilling insights humans can then analyze using System 2 thinking. For example, AI could summarize the possible clauses for a contract requiring suppliers to share sustainable sourcing information, and humans could select the most appropriate ones given the relationship with the partner, the background of the company, etc.
- **Strategic Planning:** AI aids strategic planning by offering simulations, forecasts, and scenario analyses. These provide the information humans need to engage in deep System 2 thinking, carefully considering various options and their long-term consequences. AI could provide "red team" scenarios if things go wrong in the supplier's relationship, helping buttress solutions.
- **Decision Support Systems:** AI is a powerful decision-support tool that provides detailed reports and data-driven recommendations. AI can give

a summary of all the status reviews regarding relationships with suppliers that are similar to the ones we are dealing with. Humans can then apply System 2 thinking to evaluate these inputs and make final decisions.

- **Framework-based reasoning:** Humans can apply theoretical constructs (e.g., frameworks) as a lens to critique information provided by AI to filter its output and guide further AI work. Importantly, humans can also lead the AI to use specific human-made frameworks to guide its reasoning, hence incorporating symbolic thinking derived from human research. For instance, AI can be asked to look at options with specific lenses (e.g., "triple bottom line" in the case of sustainable sourcing).
- **General critique**: GenAI makes mistakes, and human critique and quality control are valuable - even just in the form of requiring other, unrelated, and possibly specialized models to double-check the initial model output.

We aren't just talking about user interfaces. We are talking about designing a more deliberate synergy *process*, one where humans are supported holistically (UI, UX, AI itself) in their role as critical thinkers, for instance, asking us questions, guiding us through a problem-solving flow or using us to improve quality control, among others. And crucially, doing so not just one-to-one but also in groups and networks of people.

This is what I call ACI (augmented collective intelligence) System 1/2.

As it often happens with generative AI, there was a recent turn of events that might change things: the launch of more powerful reasoning models, like OpenAI o1. While these are early days, the new models show that incorporating typical human "system 2" thinking methods helps the AI achieve more sophisticated reasoning, planning, and complex problem-solving. This puts the threshold for humans higher, but it doesn't eliminate the value that we bring as custodians of System 2. For now, I see the new AI models as moving across the continuum between System 1 and 2 - a sort of System 1.5, to put it crudely.

Warning: Humans stay in, not on, the loop

Several significant challenges still need to be addressed despite the potential of integrating System 1 and System 2 thinking into human-AI collaboration.

1. Difficulty in Transitioning Between Systems: One of the main hurdles is that many people struggle with the handover between fast, intuitive thinking and slower, analytical reasoning, particularly when guiding machines in real-time. The smooth transition required to optimize human-AI collaboration is not an inherent skill for most individuals. This difficulty often leads to inefficiencies and errors when working with AI systems. For instance, people fall prey to biases and use cognitive shortcuts when reviewing AI's output.

2. Risk of Human Oversight: Recent research highlights a critical risk: the potential for humans to become overly reliant on AI, leading to a phenomenon often described as "falling asleep at the wheel." When humans overly trust AI to handle tasks, they may disengage from critical thinking, reducing their ability to catch mistakes or make nuanced judgments. This over-reliance can be dangerous, particularly in high-stakes environments where vigilance is crucial.

3. Lack of Awareness and Knowledge: We often need to design end-to-end processes across tools. Predictive (classic) AI, Robotic Process Automation, Business Intelligence tools, and Generative AI have their place in many processes if the flow is designed intentionally. The landscape of AI tools is vast and complex, and without proper understanding, users may misuse these tools or fail to use them to their full potential.

The Way Forward: Technology, Process, and People

Several solutions can be implemented to overcome these challenges.

Developing Hybrid Systems: Beyond improving user interfaces, it is also crucial to consider hybrid systems—combinations of AI and human input that dynamically shift between different types of processing.

- **Adaptive AI Systems:** These systems can start with fast, heuristic-based processing (similar to System 1 thinking) for routine tasks and switch to more complex, deliberate processing (akin to System 2 thinking) as the task complexity increases. For instance, an AI system could use quick heuristics to filter and sort data but switch to advanced (and different) models when deeper analysis (and its different flow of prompting and agents across multiple inference cycles) is required. This adaptive approach allows for greater flexibility and efficiency in human-AI collaboration.

Building Scaffoldings for Human-Machine Interaction: Creating supportive structures or frameworks is essential for smoother interaction between humans and machines.

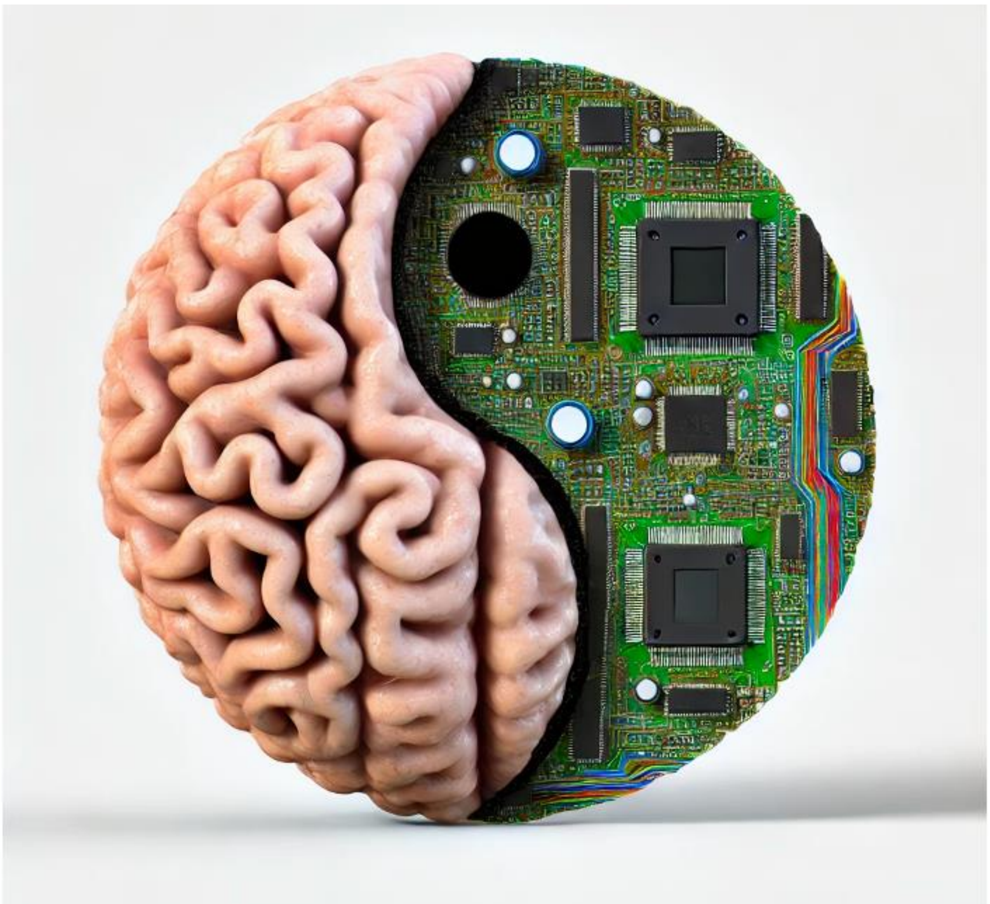
- **User Interfaces:** One practical solution our team at MIT is exploring is designing user interfaces and workflows that allow seamless switching between quick, automated responses and more detailed, analytical discussions. This interface empowers users to leverage System 1 and System 2 thinking as needed, promoting a balanced approach to decision-making.

- **Transparency and Explainability:** AI systems must offer transparent and explainable outputs, especially when engaging in System 2 tasks. When users understand the reasoning behind AI's recommendations, they can more effectively apply their analytical skills, enhancing their trust in the system and ability to make informed decisions.

Building Skills for Augmented Thinking: To capitalize on these advanced systems, we must invest in developing the skills needed for augmented thinking.

- **Human roles** will shift towards orchestration, strategy, and critical decision-making, with machines handling much of the "how" work. Key skills include critical thinking, people management, system thinking, digital literacy, and domain expertise.
- **A new curriculum** is needed to prepare individuals and teams for effective collaboration with AI, integrating foundational thinking skills, cognitive flexibility, and adaptive learning to enhance individual and collective intelligence in complex environments.
- **Humans must direct the collective cognitive attention** to the right things - the right "*whys*." We must ensure the approach is right - the right "*what*." We must critique the "*how*" that machines will increasingly suggest.

Humans, in groups and individually, are GenAI's System 2, at least for now. Let's organize ourselves accordingly.

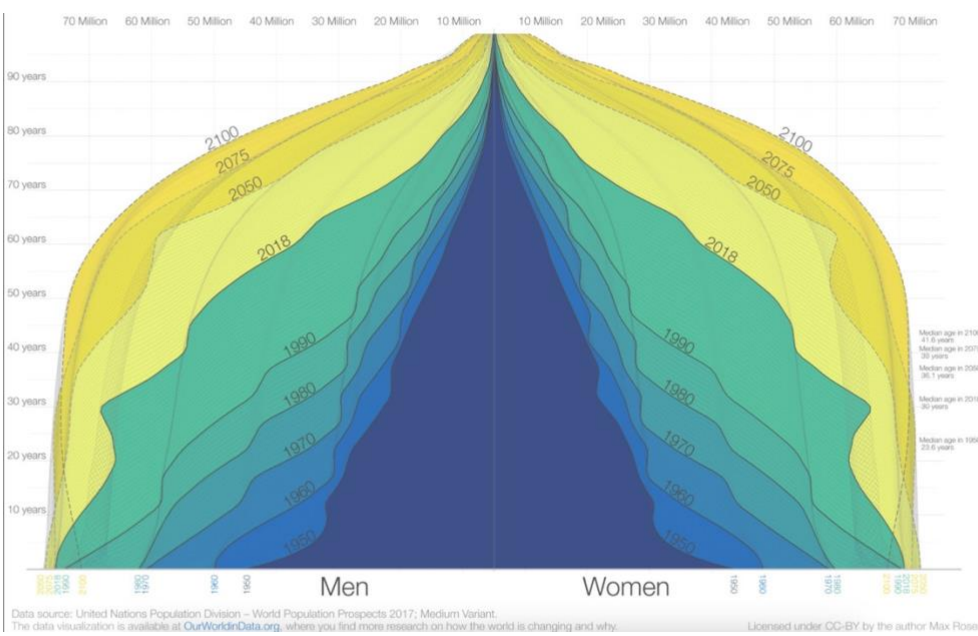


Our Collective Brain Is Ageing. What Does It Mean For Our Civilization?

An older world can be good, but only if we make it so.
Originally published in July 2024 on Azeem Azhar's Exponential View

The world is heading into a future with an increasing number of older people. The impact on economies can be devastating[1], precisely when we need resources to address large challenges such as climate change.

Is demography destiny, as some (quite a few) suggest? Are our demographics shifting from the "progress pyramids" to "domes of doom"? Are fertility policies (whose impact is debatable) the only way to address this problem?



Source: Our World in Data

I will examine the problem through a lens of collectively intelligent systems, and highlight potential challenges and solutions.

Our civilization's collective brain—one instance of what Thomas Malone termed a "supermind"—relies, in the words of sociobiologist Edward Wilson, on the interplay of three elements: Paleolithic brains with their inherent emotions, medieval institutions, and godlike technology. I want us to look at how the collective brain could change in an ageing world given that...

1. **Our individual brains age** and consequently their capabilities and incentives change,
2. **Many of our institutions' evolution doesn't keep pace** and is misaligned with the majority of citizens and people (though possibly not the short-term expressed will of the majority of the voters or believers),
3. **And our technological innovation (AI and others) has the potential to improve** the efficient frontier and total factor productivity. This would bring

economic prosperity, but their penetration and impact on established economic systems are still partially unpredictable.

By examining this interplay of challenges, capabilities, and incentives—both individual and collective—we can better understand how our ageing world might reshape our collective brain, and what to do about it.

Individual brains

As individuals age, their cognitive abilities change and in some areas degrade, especially with regard to the absorption and processing of new knowledge.

Source: Psychological Science

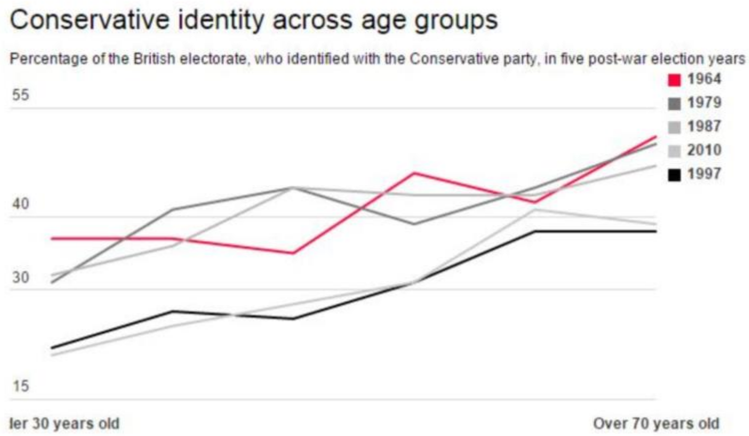
Granted, ageing today is cognitively different compared to the past, as people stay fit for longer. And yet, given the longer lifespan and the incidence of age-related conditions like dementia, we will see an unprecedented proportion of the total population with somewhat degraded cognitive abilities. Longer lifespans spent in retirement aren't conducive to people consistently keeping their minds challenged and keeping cognitive decline at bay.

The rise and fall of different cognitive abilities across the lifespan

STM = short-term memory; WM = working memory; WAIS = Wechsler Adult Intelligence Scale.

Psychological Science

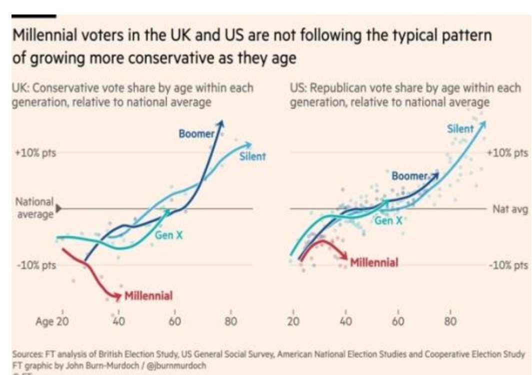
The conservative tide



Ageing societies tend to shift towards more conservative positions, especially during economic and social shocks. This rightward drift would be balanced in a typical democratic system if younger generations participated in public governance as actively as older cohorts. However, youth underrepresentation in politics amplifies the conservative shift. For instance, in the US, both politicians and voters skew older, influencing even presidential debates. Many older politicians' core values were shaped during their formative years in the 1960s and 1970s. As a result, they sometimes focus on battles that originated in that era, potentially overlooking more current concerns. This fuels the quest for finding enemies and fighting wars that are no longer a priority, and are framed in ways that aren't contemporary anymore (one can see geopolitics and identity politics through that lens, too). European politicians are getting slightly younger, but young voters are still underrepresented on the ballot, and their voice is heard less. Since younger people typically do not engage as actively in these processes, and their networks (which are often critical for nominations to important roles) are comparatively underdeveloped, there is a risk that governance becomes more conservative than the society as a whole.

Source: World Economic Forum

Recent studies indicate that Millennials in the US and UK are not following the traditional pattern of becoming more conservative with age. However, due to the current demographic structure in developed economies, where older generations still outnumber younger ones, Millennials may not have sufficient influence to counteract the overall conservative shift in societal values for some time.



Source: *Financial Times*

This effect is mediated by institutions, democratic or not.

As traditional liberal parties struggle to adapt to changing social narratives, some fringe political views may gain traction, partly amplified by social media echo

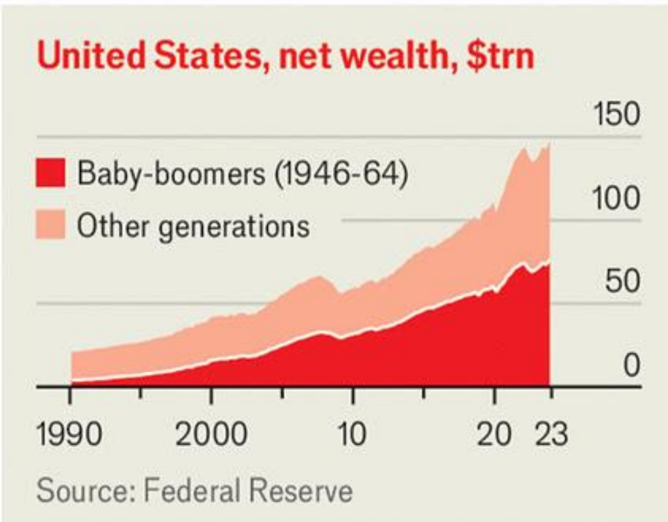


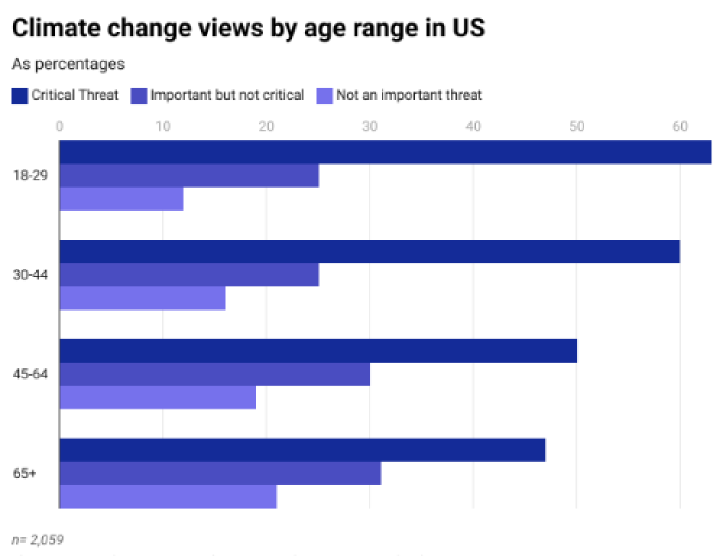
CHART: THE ECONOMIST

chambers. This polarization can increase societal tensions and conflict. And while some research indicates that older societies tend to wage war less because their ability to deploy troops is diminished, demographics could lead to perverse short-term dynamics in countries on the verge of ageing (Russia being one), and still leave the door open to autonomous-weapon warfare.

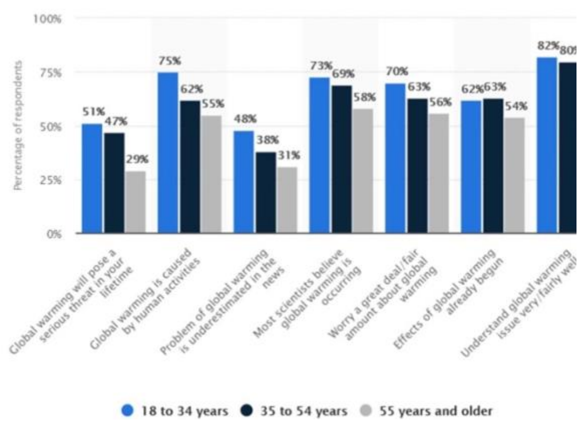
Impact on the markets

The markets, with their complex, decentralized, and dynamic decision-making processes for the allocation of resources, are another part of our collective brain. Organisational leadership has aged, for instance. Ageing societies also experience shifts in money, both private and public. As an example, housing is one the largest budget items in people's lives, but real estate taxation, and zoning laws, protect *current* homeowners (who skew older), contributing to inequalities. There is also the potential for stock market behaviour to change over time as older generations, who own a significant amount of wealth, burn through their savings - though the net effect is not fully clear yet.

And naturally, money will be needed for the climate and energy transitions. According to the IMF, as noted by *The Economist*, "rich countries will spend



21% of GDP a year on old folk by 2050, up from 16% in 2015. A quarter of that will go on pensions. The rest will be required for health- and social care provisions". Here, once more, the generational divide that drives policymaking is visible, and not just led by hard science and economics. The example below is from the US and from the UK.

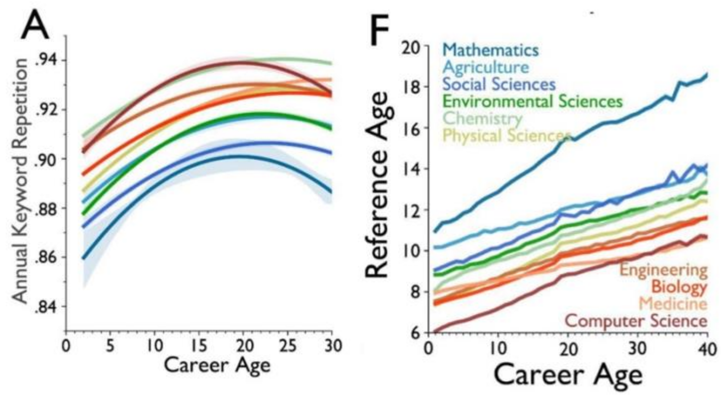


Source: Aviva, 4,000 UK adults, 2020

The perspective of European and Asian countries is less polarized, but the trend is the same, especially when it comes to the impetus for action.

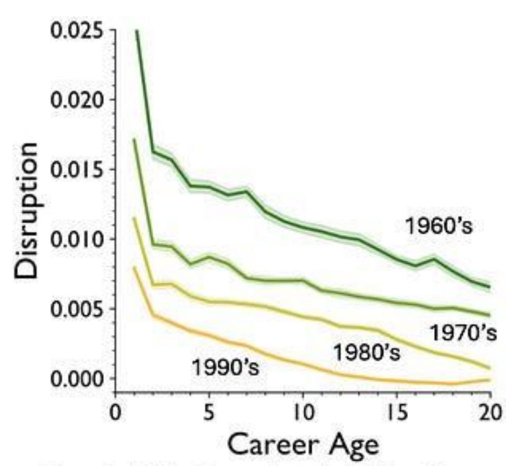
Innovation risk

Finally, and crucially, ageing societies can see a change in how they innovate. Starting with (ageing) academic leadership - crucial for inventions - where incentives are stacked in favor of academics going deeper and for more years into their narrow field rather than looking across disciplines. The charts below illustrate for instance how papers referenced by older academics are on average older and their research is less likely to disrupt the state of science and more likely to criticise emerging work.



From Cui, Wu, and Evans (2022)

This, combined with the burden of knowledge and specialisation, can slow down disruptive innovation and breakthroughs. The chart below shows the effect of age on the novelty of academic research over the life of researchers, and over time.

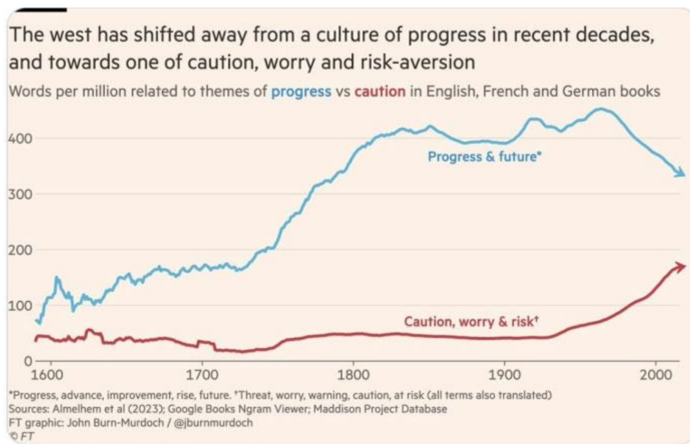


From Cui, Wu, Evans (2022) via Matt Clancy

Two famous quotes encapsulate the potential problem. Arthur C. Clarke quipped that "[w]hen a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong". Max Planck famously lamented that science often shifts paradigms only when old academics die.[2]

Innovation is not just invention though. An older population might mean slower uptake of new ways of doing important things. Landmass use for instance will be crucial in our climate transition, but in Europe, the region where environmental policies are politically least controversial, the average farmer's age is about 60 (and only 10% are below 40). More money for the older people might mean less money for the young, their education, and the support they need to credibly enter the workforce and change work practices, with a potential impact on the speed of adoption of innovation. In another example, energy and transportation senior executives in incumbent Western companies might feel that, if they delay things just enough, they might be able to juice the previous investments in older technology (fossil fuel generation or ICE, for instance) and retire without needing to push through hard changes that could jeopardize their financial profits in the short term. The possible onslaught brought about by Chinese EV companies, whose economies of scale and Wright curves have driven cost reduction can now only be fought with import tariffs, shows the risk of misreading the time it takes to realize the benefits of innovation. In general, the average age and time-to-retirement of senior executives might skew decision-making in companies, especially publicly listed ones, although the effect may not be linear and some older leadership teams might indeed focus on leaving a lasting legacy.

Trying new things also requires a certain amount of risk-taking. Judging from the



data in the chart below, and despite all caveats required in such analyses, there is reason to believe that ageing, wealthier societies see progress as more of a half-empty glass than they did in the past.

Clearly, we cannot allow a small minority of people, who have the requisite tech capabilities and skew much younger, to work alone on significant technological innovations with the potential to create significant risk to everyone. There's more than a grain of truth in the claim that Silicon Valley's youthful (immature?) ethics is an insufficient moral compass in these times. But the option of stifling the right type of innovation is not viable either. An older population might lose touch with the younger minority able to drive innovation fast, which would be a dangerous mistake.

Bridging the gap

Here are several ideas to address the impact of an ageing society on our collective intelligence. There are surely many more, and doing them justice would require much more depth than what we can discuss here.

From a resource standpoint, we will need to do things like

- (a) **Do more with fewer workers**, and/or fewer workers in "their prime", which means better productivity for workers in all age brackets; this can be achieved by augmenting people's abilities, slashing non-value-added work, and attacking inefficient ways of working (think: bureaucracy paper-based work),
- (b) **Get more people into the workforce**, and give them a solid chance. Think, people out of work, including senior and female participation incentives and corresponding jobs design; fixing the skill mismatch by providing better signals to young people as they train through universities, and do a better job at targeted, continuous education; fixing location mismatch by improving immigration flows, and increasing remote work across all sectors.

AI can certainly support these. While most of the limelight is taken by exciting technological progress, a large part of the battle is being fought on the front of organizational design, including processes and people's skills.

Consider the following. The promise of AI is an explosion of new business and operating models that were impossible before, including all sorts of productive augmentation to scarce workers (think of the lack of skilled tradespeople) or workers whose physical and mental conditions may need support. But today's institutions, their processes, and our culture - and even our technology - aren't necessarily set up to leverage people beyond a certain age.

The current volatile and fast-evolving conditions make many older people feel exposed, and give them even more incentives to barricade themselves while they can - and who wouldn't. Just look at South Korea, one of the fastest-ageing countries yet, lacking a public pension scheme for many, and where 40% of elderly are under the poverty line? And gig work isn't likely to solve the problem by itself, especially in its current form.

We must make the future of work more elderly-friendly, just not at the expense of younger people. For instance, AI-supported coaching of people to better collaborate across ages (and cultures), or smarter job design, recruitment and training, will go a long way. Retirement should also be reinvented, for instance with useful jobs helping keep individuals economically independent, socially connected and mentally fit.[3] Older people have many social skills that can prove invaluable to individuals in need - children and adults - and their communities. Keeping them mentally and socially engaged improves their mental abilities, and keeps them fitter overall, which also helps keep the related healthcare costs down. AI, done well as part of human-centered processes, can help here - augmenting the older professionals' individual intelligence, and their ability to work in groups with others.

On the (unsustainable) cost side, think of the future of healthcare whose cost is disproportionally allocated to older generations, where artificial intelligence can tackle not just the problem of treating illnesses, but also managing wellness, which is a much less well-monetized space and attracts a minority of the investment compared to clinical health treatments. Or think of the revolution in providing personal tutors to learners of any age, combating Baumol's disease reflected in the constant increase of the cost of education. This, for instance, would help not just students but anyone who needs to be redeployed and learn new skills (as AI takes over parts of jobs, or as the climate transition shifts economic activity) or even a new culture (supporting, for instance, scaled-up immigration).

In general, it seems clear that the revolution in generation and access to knowledge can help solve many problems, assuming the right incentives are in place for individuals and institutions - which brings us to the last point.

From a governance standpoint, we will need to (c) **redesign our collective governance systems** to give younger people more of a voice. A digital evolution of the governance structures could yield new means of voting and civil engagement, as pioneered in Taiwan. Governance needs to reflect the changing demographics, with a 16-year-old with 80 years to live able to vote on environmental policies, for example (some EU countries allow that already) - considering that octogenarians with few years of life left are allowed to. More futuristically, it is not impossible to think about younger generations having more sway in future-endangering policies (remaining-lifespan weightage?), or being better represented by AI agents who help their networks organize more effectively, and systematically but democratically pressurize authorities.

There are many more general interventions that we could design at a systemic level, that we don't have space to discuss here (but are discussed in depth here).

An older world can be good, but only if we make it so

We don't have the option not to try. Our ageing society breeds conservatism - writ large. But conservatism clashes with a world that has already integrated (and priced) future expectations into its present systems—be they the stock market, pensions, or general welfare. We're essentially borrowing from the future, banking on growth and improvements in efficiency, even if those innovations are yet to be realized. However, the current demographic trends undermine these assumptions, and as a result, we cannot allow demographics to stifle our focus and drive for progress. While "steady state" may serve well to some as a political slogan, it is economically unsustainable, and consequently socially untenable. When financial shortfalls and lack of opportunities arise, they breed conflict, leading to significant, perhaps catastrophic, disruptions.

Fertility incentives are unlikely to give us the demographics that we want. And we can't pretend that our ways of working, governing, and innovating are "just fine" in the presence of such a large demographic shift. Conversely, managing a shrinking population well could be a lever for reducing humans' footprint on the planet without compromising our welfare.

While this may not make for easy conversation at the kitchen table - or in parliament - the upside is huge. And the longer-term downside, if we don't deal with it with all the tools we have, is very unpalatable.

Cut Climate Invention-to-Innovation Time

The cycle of invention (idea successfully prototyped) to innovation-at-scale (widespread implementation of the new practices) typically takes decades. For climate change, we just don't have that time. Yet, we can intentionally compress it with existing technologies and organizational design.

Every science—and ultimately technology-based revolution—typically takes decades to percolate deeply into the world because of the slow process of "learning by doing." This process often starts with academia, with some R&D pilots successfully executed, then the most innovative managers adopting the new practices, and finally , most others following many years later.

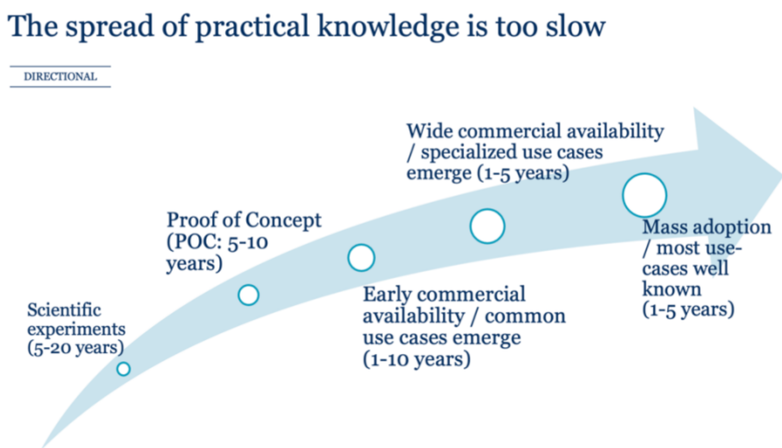
This is innovation's death by thousands of small cuts. One of the most striking examples of that problem is the slow progress in evolving healthcare systems worldwide, with its immense variance in the deployment of tried-and-true practices (e.g., India's Aravind eye care process for cataract treatment, with its order-of-magnitude cost improvement at comparable quality, which, a decade after scale, is still not widely adopted worldwide).

We can't rely on established knowledge-transmission mechanisms

For climate change, the "typical cycle" is not nearly good enough. We can't wait for five years until heat-pump installation capacity ramps up; we can't wait for established regenerative agriculture practices for specific microclimates to spread to enough farmers who don't speak English well; we can't wait for enough municipal utilities to learn how to incentivize and enable citizens for efficient energy usage; we can't wait for a serendipitous uptake in the long tail of cities, regions, and countries that are not exposed to the most recent technology, methods, and applications.

Many strictures exist: from upstream scientific to downstream practitioners access to knowledge repositories is not always as easy as it should be (academia and media); many professionals don't know how to thoroughly harness social media where new ideas surface; language barriers make it hard for the "global South", among others, to access and share new things. And the natural tendency of experts to silo their knowledge and try to find the next big thing in their field, whereas we know that innovation comes from the combination of existing ideas.

Industry and generally internet media are also not doing that job well enough. Thanks to algorithms tuned to maximize advertising and stickiness, meme-able noise often obfuscates the signal, and



finding what's relevant is still too hard or expensive (e.g., paywalled content).

So despite the excitement about climate startup funding and corporate net-zero commitments, at least one aspect remains seriously neglected: the intentional crystallization and sharing of practical, specialized, knowledge so it productively "touches the ground" and can be recombined with existing ideas, processes, operations etc. That's a clear multiplier of impact but unsexy for many entrepreneurs and investors, and often left to either individual firms' marketing, or to well-intentioned but under-resourced NGOs and other public institutions - including educational ones - that struggle with both granularity of information and speed of change. The outcome is a frequent reinvention of wheels.

We can do better today

This is not just about media or training. Both help, but in isolation, and when executed in a traditional manner, they have significant limitations. What works is a new organization for the *knowledge of networks of people augmented by intelligent technology*: **Augmented Collective Intelligence**.

Today we have access to methods for knowledge formalization, retrieval, and sharing, vastly superior compared to the past. Google, Wikipedia, and the Web2 revolution (from WordPress blogs to Reddit, LinkedIn, Substack, Medium, etc.) have shown potential; yet they're not yet "finishing the job" of making relevant and practical climate-change information efficiently available to most relevant people. A minority of experts and practitioners know many information sources and can monitor them efficiently, but most others can't. That's significant leakage in the invention-to-innovation cycle. We can do better.

The table below summarizes the main idea. Hyperspecialized collective-intelligence "utilities" could accelerate the spread of high-momentum/low-signal *content* (both practical enablement and broader learning), and support the identification and engagement of relevant *people* (experts and practitioners). These infrastructures can use new natural language capabilities, and build *knowledge graphs* that facilitate two crucial processes: first, finding and combining granular information, i.e. the "what" (e.g., new ways of implementing heat pumps cost-effectively in areas where energy is expensive and unreliable); and second, pinpointing experts, i.e. the "who" (e.g., people or organizations who have codified the respective processes and can help on the ground).

The spread of practical knowledge can greatly accelerate DIRECTIONAL

What	Who		Knowledge tools		Speed	
	Supply	Demand	From	To	From	To
Scientific experiments	Science labs / R&D	R&D	Scientific papers, some research portals	Addition of explicit knowledge graphs enables combinatorial innovation, identification of players	5-20 years	TBD
POC	Translational R&D, pioneers	R&D	R&D, science papers, whitepapers		5-10 years	TBD
Early commercial availability / common use cases emerge	Startups, specialized corporates	Innovation dept. of pioneering firms	Enterprise thought leadership; some training from suppliers; some press; specialized social networks; academic courses	Hyperspecialized "collective-intelligence utilities" accelerate the spread of high-momentum (and low-signal) knowledge (both enablement and learning), and more effective allocation of relevant experts and practitioners	1-10 years	50% less
Wide commercial availability / specialized use cases emerge	Scaled up new players, corporates	Early majority's operations dept.	Marketing; training (incl. enterprise and early vocational training); social networks; press		1-5 years	90% less
Mass adoption / most use-cases well known	Large new players, corporates, disruptors	Mainstream operations dept.	Marketing; commoditized training; vocational learning; social media	Explicit knowledge graphs facilitate identification and allocation of practitioners	1-5 years	TBD

Time

CORE IMPACT HERE

The uptake would be that the new granular, practically implementable knowledge could now reach not just the pioneers or the "hackers", but also mainstream professionals open to new ideas. That is the *early majority* of users.

There isn't a clearly defined category for this type of work. It sits between social media, professional networks, education, training, open-source solutions, and even thought-leadership marketing. But the **building blocks already exist**. For instance, Microsoft has Viva, LinkedIn, and Bing, which - combined - potentially have the full solution both within and *outside* of organizations. Others could use off-the-shelf tools that combine content and social media scrapers, perhaps using additional sources such as Google's open-source science and data repositories, or the amazing G-DELT machine-translated world news, or interesting new tools like Diffbot. Climate solutions startups like Ubuntoo (disclosure: I am an advisor there) already curate knowledge for innovation. Content providers, from scientific journals to Twitter, Reddit, and Quora, could make it easier to access rich APIs for this.

The sharing and combination of the world's relevant collective knowledge can be intentionally engineered thanks to new digital technology and practices. We could soon live in a world where detailed, specialized *"how-to"* knowledge for climate mitigation and adaptation is available on a browser that millions of people can readily access. Then, a broader base of people will have a fighting chance to tackle the most significant challenge humanity has ever faced.

Let’s ignite thousands of climate “superminds” powered by a shared infrastructure.

If The World Knew What the World Knows

Genius and stupidity seem to coexist at an unprecedented scale in our world. As Edward Olsen said, the interplay between our Paleolithic brain, medieval institutions, and advanced technology is at the root of many of our struggles. The collective intelligence emerging from those three elements is constantly tested and often fails—from populism to social media gone awry to pandemic unpreparedness and climate change. It often feels like we are fighting tomorrow's challenges with yesterday's intelligence.

But there's one significant reason for optimism, as one *very* large resource is largely untapped. *Our world routinely throws away or ignores the knowledge we create*. You can see it in your own daily work, and the work of your organizations: every day, we reinvent wheels, and we don't access the right people (or organizations) at the right time to find (or remember) solutions. Our collective brain isn't functioning as well as it could.

An infinite engine of knowledge

In the last twenty years, thanks to the web, we have wired our collective brain in unimaginable ways. The world creates an astonishing amount of data and knowledge and makes it available online. It connects people in incredible ways that would have felt like sci-fi at the turn of the millennium. (Hundreds of examples of organizations, movements, and building blocks that harness this power have been inventoried.)

Yet, when it comes to harnessing planetary knowledge, we haven't seen anything yet. Today there's immense and untapped potential because of the convergence of a few powerful vectors. Consider these examples.

One of the most important innovations of the last twenty years has been the **search engine**, which Google describes as intended to "organize the world's knowledge." Today, even video and audio content can be easily searched.

AI's natural-language models have enormously progressed in the last years, leading to astonishing tools such as GPT-3 and its successors, which have made language understanding and generation a lot easier. Beyond written language, image processing and generation like Dall-E2 or Stable Diffusion have also evolved by leaps and bounds (as a presage of future things, Stable Diffusion's Text-to-Image prompt result database is now being mined by a dedicated search algorithm). By the way, this is not an unalloyed good, and needs careful design and implementation, as Meta found out in their recent experiment.

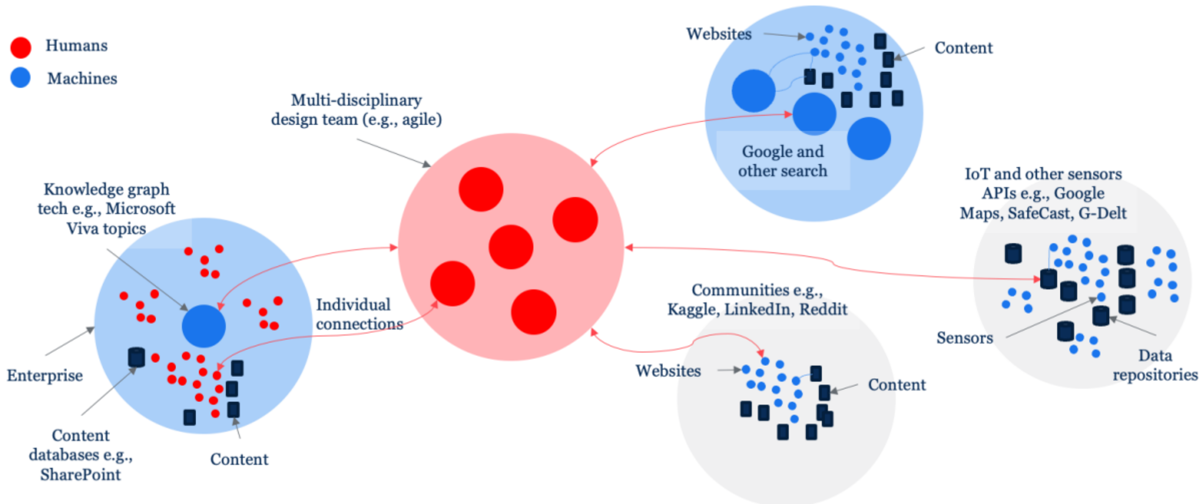
Knowledge-graph technologies that establish relationships between concepts, people, and organizations (“entities”), make the world’s knowledge even easier to mine, especially when combined with large language models (LLPs). New tools use that to enable richer search (and this) and, when combined with natural language understanding (for instance, in science, this, this, this, this, this, this, and this), hold promise for the exploration of specific topics (e.g., this, this, this, this, this and this). Knowledge graphs may soon be fed by AI natural-language technology, industrializing the organization of richer information (imagine mining the relationships between drugs, genes, and proteins, evinced from scientific texts). And the ongoing re-mix of everything made through social media makes connections between ideas, people, and organizations explicit — some of which can be mined through publicly accessible APIs.

Data science, including its crowdsourced citizen data-science form, enables the use of new and existing sources of data, including the increasing amount produced by the Internet of Things (IoT), which is both public (e.g., heat measurement), private (e.g., Google’s land development tracker), and crowd-based (e.g., Arduino-based sensors).

People simply share more: thanks to self-publishing tools, and because of the importance of enterprise and personal thought leadership, the web is awash with publicly available content from companies that would have been considered trade secrets only a couple of decades ago. Scientific knowledge is increasingly retrievable, through specialized search engines (e.g. Google Scholar), portals, and networks (e.g. Researchgate), and because of the mounting pressure to make it freely accessible.

And it is not just “asynchronous” knowledge access. Modern cloud technology and sophisticated data compression algorithms make **video and voice** connectivity ubiquitous at increasingly low data speeds, making synchronous knowledge retrieval and generation more frictionless than ever.

As a result, augmented collective intelligence (MIT’s Malone calls them superminds) is emerging, as the image below illustrates for the example of enterprise innovation teams, who harness the internal and external ecosystem as an extension of their own brain.



Still nowhere near our organic counterparts

But our collective technology and methods pale in comparison with what happens elsewhere. The world does countless "natural" experiments (both in our society as well as in nature) that aren't harvested— unlike the "active inference" that our brain and in a way the natural world do. Take the following examples:

Search engines' algorithms, and their use, are still largely driven by advertising markets, not knowledge industries. Research points out that AI can give innovators superpowers by, among others, improving search for knowledge across domains (if we enable people with the cross-disciplinary skills that they need to make combinatorial innovation happen). But search engines and commonly-used methods do not make truly advanced search available to most people. For instance, they don't explicitly allow the exhaustive visualization of knowledge graphs, so that one could identify both content and people (and organizations) — as well as explore adjacent fields. Not all meaningful websites and content are inventoried. Much of the "new" action currently remains *within* enterprises through machine-learning-based knowledge management (such as Microsoft Viva Topics), but the overall knowledge ecosystem is many orders of magnitude larger.

Social media algorithms' recommendations optimize for predicted engagement (e.g., likes, or shares), not problem-solving or creativity. And try to follow the right people and the right topics isn't effortless: one can't easily find people to follow based on the field they're competent in. Similarly, **professional social networks** such as LinkedIn are not optimized for skill-based search ("which people work in my field?") and do not facilitate field exploration ("which subfields exist, and who works there?”) or validation of ideas (e.g., assessing people’s claims credibility by checking their — or their network’s — skills).

Natural language models could proactively propose novel combinations of concepts for humans to refine, but they have not yet been used for that purpose.

Data science, and translating science into respective models, is still an elite job. However, data crowdsourcing (e.g., citizen science), and increasingly easy-to-use tooling (e.g., XGBoost) show that the barrier to entry can be further lowered.

Much knowledge, especially publicly-funded research, still sits **behind paywalls**, preventing deep mapping and access.

Too many people are so specialized that they **can't combine knowledge** from different domains to unlock combinatorial innovation, and training is often focused on specialization instead of so-called "T-shaping."

Surprisingly, **language barriers** are still significant and end up siloing up the world’s knowledge. Think about it: web searches only show results for same-language sites: if you are in the US and look for “heat pump installation methods”, you typically won’t see content from (machine-translated) German, Japanese, or Chinese sources. And while translation engines like Google Translate have improved remarkably, they’re still not used pervasively yet in a range of potential knowledge-sharing applications.

As a result, *we collectively don’t learn enough from the experiments made elsewhere*. Think of “Global South” practitioners quickly learning from cost-effective climate adaptation projects in other countries, irrespective of whether they are documented in Indonesian Bhasa, Spanish, Urdu, Swahili, Hindi, or Chinese. Conversely, developed countries' practitioners fail to access sources of “reverse innovation” — lower-cost ideas developed under significant budget constraints. And, generally, knowledge “backwaters” exist: users in many (non-English speaking) countries prolong the use of old knowledge because they don’t have access to the right networks in real-time (think of old schoolbooks and non-English language internet pages for technical topics).

Sadly, our **organizational design** practices reflect the issue: strategic knowledge creation and management isn’t a C-suite role, and that job is often fragmented across departments — domain practice groups, the CIO, sales support, etc. which weakens the much-needed enterprise transformation. Across even broader ecosystems, incentive systems are still broken, as attested by academia’s struggles to give appropriate credit and encourage more creative exploration.

Finally, and ironically, the respective digital product ecosystem doesn't attract as much attention and investment as others (venture capital, anyone?).

What we need to do

I have argued elsewhere that in order to amplify and accelerate innovation cycles, we need to build "supermind utilities" — possibly as public or partially open-sourced goods so that the global community can access them. They could be financed by governments, private individuals, or corporations. Over time, the return on such investments will attract more private capital, crowdsource contributions, and help develop business models that eschew advertising and make money by stimulating our pre-frontal cortex, not our amygdala. (The potential promise of some web3 technology could help, as and when it gets out of its current hype and greed cycle.)

To be clear — there's likely a solid business case to build commercially viable digital products that cater to a type of "knowledge super users". The current challenge is to show them (and their C-suite), an easy and exactly quantifiable return on investment. As is often the case, the most sophisticated users, and the companies with the most foresight, will lead the pack.

Over 2,500 years ago the library of Alexandria ignited innovation across a chunk of the ancient world, and innumerable efforts have helped build repositories of knowledge over the centuries. The word "university" originally meant "community", and universities received funding to strengthen those (analog, organic) superminds — helping the respective networks and their knowledge converge. In the 21st century, augmenting the world's collective intelligence by building such knowledge utilities sounds like a reasonable thing to do.

These superminds will generate a superior intelligence, emerging from the network of knowledge and skills that exists below today's comparably superficial web-based interactions. They will help us fight tomorrow's challenges with tomorrow's intelligence, across:

- **known-knowns:** problems whose solution exists elsewhere, so collectively remembering and learning what works,
- **known-*un*knowns**, by creating and deciding on solutions that we struggle with, and
- ***un*known-*un*knowns**, by sensing low-signal but high-momentum trends that could quickly turn into major opportunities or threats.

Of course, lots can go wrong. To start with, we will need ways to mitigate our collective tendency to fall for unsubstantiated claims, counter rogue actors, and generally reduce trolling and abuse. But with the right incentives, methods, and capability, it sounds plausible that we will be able to emulate, for instance, Wikipedia and its collectively-enforced quality control.

Every single hour, the Earth receives from the Sun the amount of energy that the entire human civilization consumes in a year. We are getting better at harvesting that power. There is reason to believe that we are "leaving knowledge on the table" in similar proportions, and by harnessing our collective knowledge, we could harvest our collective *cognitive* power.

Much of our innovation challenges are addressed as "design" problems: typically tackled by small groups of experts, with comparatively limited access to the world's collective intelligence. Instead, we can turn them into "search" problems, which makes them likely to be tackled by the many experiments that happen in the world every day and are documented in an increasingly large and accessible knowledge corpus.

Building on today's technologies and methods, there's much that we can do about it. Let's solve tomorrow's problems with tomorrow's intelligence. **Let's go build** superminds**.** 🧠

How to Build Superminds

These perspectives have and can be implemented today. Contact us if you want to use AI-Augmented Collective Intelligence to design and build the next generation of what your organization and teams do.

At supermind.design you will find resources, from workbooks and videos to a database and interactive AI tools.

Let’s go build superminds.

Download the latest version at **www.supermind.design**/resources

Contact: info@supermind.design
https://www.linkedin.com/in/ggiacomelli/