The Chinese Room
Important Concepts
Cognitive Science is an interdisciplinary approach to the study of the mind and its functions.

It usually consists of philosophy, psychology, neuroscience, linguistics, anthropology, computer science and artificial intelligence, but can include other disciplines.
Computational Theory of Mind is an umbrella term for a family of views that hold the view that mental operations are *computations*. 

brain $\approx$ computer (an information processing system)

cognitive capacities $\approx$ programs
A module is an innate neural structure which has a distinct, evolutionarily-developed function. In other words, it is a “program” that performs some cognitive function.
Weak AI is a form of artificial intelligence that can:
A. perform narrowly-defined tasks, e.g., data collection, speech recognition, driving, etc.; but
B. is not conscious (or “self-aware”).
Strong AI is a form of artificial intelligence that can:
A. perform narrowly-defined tasks, but also
B. displays general problem-solving skills, since it is
C. conscious (or “self-aware”).
Question: What is Artificial Intelligence?
Various theorists lament over the “moving goal posts” for what counts as artificial intelligence (e.g., Hofstadter 1979: 26).

The basic complaint is that once some particular task which appears to be sufficient for intelligence (of some sort) has been mastered by an AI, then that particular task is no longer sufficient for intelligence.
For his part, Yudkowsky (2008: 311) argues that artificial intelligence refers to “a vastly greater space of possibilities than does the term Homo sapiens. When we talk about ‘AIs’ we are really talking about minds-in-general, or optimization processes in general” (italics in original, emphasis added).
Here are some potential disruptions that could be brought on by AI...
Act I: Human Obsolescence
Perhaps the best case scenario is the building of an innocuous (or “friendly”) superintelligent AI that can successfully solve all human organizational problems and blaze forward on all technological matters.
Space Colonization
As it turns out, there are various voluntary actions that add to the total level of happiness of a particular person, e.g., steering clear of constant traffic noise, less stressful commutes, and, relevant to us, avoiding the feeling of lacking control (see Haidt 2006, chapter 5).
In a classic study, David Glass and Jerome Singer (1973) exposed two groups of subjects to loud bursts of random noise. Subjects in one group were told they could terminate the noise by pressing a button, but they were asked not to press the button unless it was absolutely necessary.
In the second part of the experiment, the subjects who thought they had control were more persistent when working on difficult puzzles, but the subjects who had experienced noise without control gave up more easily.
In a post-scarcity economy, what will be left for humans to do?

In what activities will humans find meaning?

See Danaher (2019) for a potential solution...
Act II: Automation
Example #1
Unemployment and Social Unrest
According to a recent study, about 47% of US employment is at risk of being robotized (Frey & Osborne 2013).
Overreacting?!?

Another (UK) study says it’s only 38%...

And other researchers (from RAND) are more worried about A.I. starting nuclear war by 2040.
Storytime!
Symbolic artificial intelligence is an umbrella term that captures all the methods in artificial intelligence research that are based on high-level "symbolic" formal procedures. It is based on the assumption that many (all?) aspects of intelligence can be achieved by the manipulation of symbols, as in first-order logic; this assumption was dubbed as the “physical symbol systems hypothesis.” Symbolic AI was the dominant paradigm of AI research from the mid-1950s until the late 1980s.
Abbildung 3: "Find a block which is taller than the one you are holding and put it into the box".

Terry Winograd and SHRDLU, 1972
Eliza is a mock Rogerian psychotherapist. The original program was described by Joseph Weizenbaum in 1966. This implementation by Norbert Landsteiner 2005.

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EE</td>
<td>LL</td>
<td>II</td>
<td>ZZ</td>
<td>AA</td>
<td>AA</td>
</tr>
<tr>
<td>EEEE</td>
<td>LL</td>
<td>II</td>
<td>ZZ</td>
<td>AA</td>
<td>AA</td>
</tr>
<tr>
<td>EE</td>
<td>LL</td>
<td>II</td>
<td>ZZ</td>
<td>AAA</td>
<td>AAA</td>
</tr>
<tr>
<td>EEEE</td>
<td>LLLLL</td>
<td>III</td>
<td>ZZZZZZZ</td>
<td>AA</td>
<td>AA</td>
</tr>
</tbody>
</table>

ELIZA: Please tell me what's been bothering you.
YOU: The weather is awful.
ELIZA: I'm not sure I understand you fully.
YOU: I hate the current weather.
ELIZA: You say you hate the current weather?
YOU: Yes, that's what I said.
ELIZA: I understand.
YOU: Great.
ELIZA: Please go on.
YOU: I would say it's your turn now!
ELIZA: Why are you concerned over my turn now?
The Chinese Room Thought-Experiment
“It seems to me obvious in the example that I do not understand a word of Chinese stories. I have inputs and outputs that are indistinguishable from those of the native Chinese speaker, and I can have any formal program you like, but I still understand nothing. Schank’s computer, for the same reasons, understands nothing of any stories, whether in Chinese, English, or whatever...” (Searle 1980: 186).
Searle is thereby denying the possibility of strong artificial intelligence under the assumption of the physical symbol system hypothesis.
“Whatever purely formal principles you put into the computer will not be sufficient for understanding, since a human will be able to follow the formal principles without understanding anything, and no reason has been offered to suppose that they are necessary or even contributory, since no reason has been given to suppose that when I understand English, I am operating with any formal program at all” (Searle 1980: 187).
1980s-ERA NEURAL NETWORK

- Input layer
- Hidden layer
- Output layer
- Node

Links carry signals from one node to another, boosting or damping them according to each link's 'weight'.

DEEP LEARNING NEURAL NETWORK

- Input layer
- Multiple hidden layers process hierarchical features
- Output layer

Identify:
- Light/dark pixel value
- Edges
- Combinations of edges
- Features

Identify combinations or features:
- ‘George’
Example #2
Over Reliance on AI

Some governments, in an effort to be ahead of competitors, begin to delegate and automate important decisions to AI.
During the process of Machine Learning and Deep Learning, we are not aware of the explicit connections and inferences being made by the AI.
“Once upon a time, the US Army wanted to use neural networks to automatically detect camouflaged enemy tanks. The researchers trained a neural net on 50 photos of camouflaged tanks in trees, and 50 photos of trees without tanks. Using standard techniques for supervised learning, the researchers trained the neural network...”
“Wisely, the researchers had originally taken 200 photos, 100 photos of tanks and 100 photos of trees. They had used only 50 of each for the training set. The researchers ran the neural network on the remaining 100 photos, and without further training the neural network classified all remaining photos correctly. Success confirmed!”
“The researchers handed the finished work to the Pentagon, which soon handed it back... It turned out that in the researchers’ dataset, photos of camouflaged tanks had been taken on cloudy days, while photos of plain forest had been taken on sunny days. The neural network had learned to distinguish cloudy days from sunny days” (Yudkowsky 2008: 321).
Example #3
Partial Automation

Success in various domains of AI might stagger and we will only partially automate many tasks.
Emily Guendelsberger (2019) gives various examples of how companies are using optimization algorithms for scheduling and micromanaging which have adverse effects on workers.

Note: Click on the image on the right for an interview of Guendelsberger.
“I was hired for picking, which is generally regarded I think as the least desirable job at warehouses. We would get a cart and we’d have the scanner. There were about, I think it was four or five steps to going out to locate the coordinates that it gave you and find the actual, whatever the thing was. You would just walk around all day and do that. Every single step of this was accompanied by a little countdown. At the bottom of the screen, there is a blue bar. It says how many seconds you have left to do it, and then it would start ticking those seconds down. So it’s kind of constantly reminding you like, ‘Hey, move. Keep moving. Keep moving. You are not keeping up’” (Guendelsberger in Intercept Interview).
Example #4
AI-Enhanced Political Advertising and Misinformation Campaigns
Successes so far:

“Mishne and Glance (2006) showed that positive sentiment is a better predictor of movie success than simple buzz (keyword) count. Liu et al. (2009) reported a sentiment model for predicting box-office revenue...

Tumasjan et al. (2010) even showed that simply part mentions on Twitter can be a good predictor of election results...

Instead of using bullish and bearish sentiments, Zhang et al. (2010) identified positive and negative moods on Twitter and used them to predict the movement of stock market indices such as the Dow Jones, S&P 500, and NASDAQ” (Liu 2015: 6-7).
Things We Know:

1. Hackers have already started to weaponize AI.
2. Some tech experts have suggested that we voluntarily discontinue research in Artificial General Intelligence.
3. Some countries (e.g., Russia) have already deployed political interference campaigns with non-negligible results.
The most alarming hypotheses, however, might be like those of philosopher Nick Bostrom (2014) who thinks that general-domain artificial intelligence will lead to an intelligence explosion that could spell the end of the human species.
“If I play chess against a stronger player, I cannot predict exactly where my opponent will move against me—if I could do that, I would necessarily be at least that strong at chess myself.

But I can predict the end result…” (Yudkowsky 2008: 320).
“Before the prospect of an intelligence explosion, we humans are like small children playing with a bomb. Such is the mismatch between the power of our plaything and the immaturity of our conduct. 

Superintelligence is a challenge for which we are not ready now and will not be ready for a long time. We have little idea when the detonation will occur, though if we hold the device to our ear we can hear a faint ticking sound” (Bostrom 2014: 319).
References

Representative Research in Social Psychology.
PHILOSOPHY OF PSYCHOLOGY, 407.