The Imitation Scam:
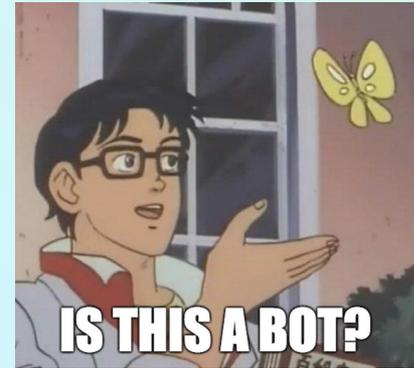
Faking Online Personalities via A.I. Synthesis of Language, Images, and Video

Scott H. Hawley, Ph.D.

IS THIS A BOT?

National Science Teachers Association, March 1, 2019

Setting up the Film: "The Imitation Game"

ALAN TURING

We're going to watch a movie about the great mathematician and computer scientist Alan Turing, The film is mostly about Turing's code-breaking work at Bletchley Park in WWII. (Note: the movie distorts much of the history.)

BLETCHLEY PARK

Was a secret location in middle England where many people worked in great isolation, in an effort to decode intercepted German messages.

It's now the home of the National Museum of Computing.

What's the Imitation Game?

After WWII, Turing's research centered on theories of computation & mind, and whether one could build a machine that could reproduce the thought processes of the human brain.
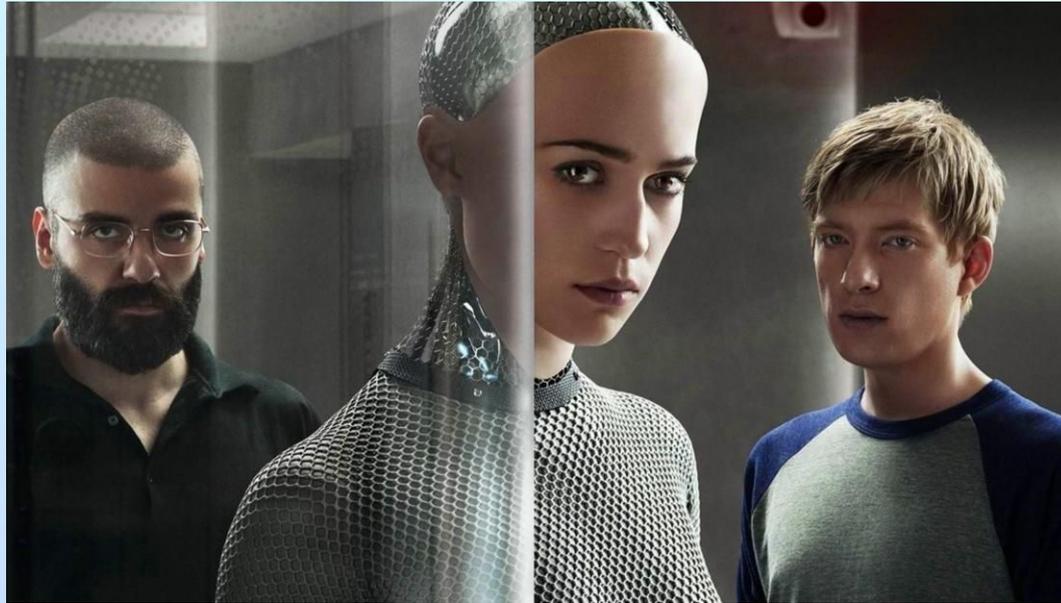

QUESTIONS

What would count as intelligence?

How would you know if a machine could think?

THE IMITATION GAME, AKA THE "TURING TEST"

Turing's answer (1950) centered on the use of **language**: By only passing messages back and forth, could you **fool a human 'player'** into thinking they're 'conversing' with another human, when in fact the responses were originating from a computer?

(Different movie)

ELIZA: the First Chatbot

1964-68: Joseph Weizenbaum created ELIZA, the first chatbot, as a Rogerian psychotherapist. It was just doing pattern-matching....

Later Weizenbaum: big into AI ethics, e.g., that we absolutely should not have AI therapists!

Does the program really understand?

LANGUAGE WITHOUT COMPREHENSION

1980's: John Searle proposed "**Chinese Room**" thought experiment: Even if it didn't 'know' Chinese, a computer could just look up responses in a (long) rule book!
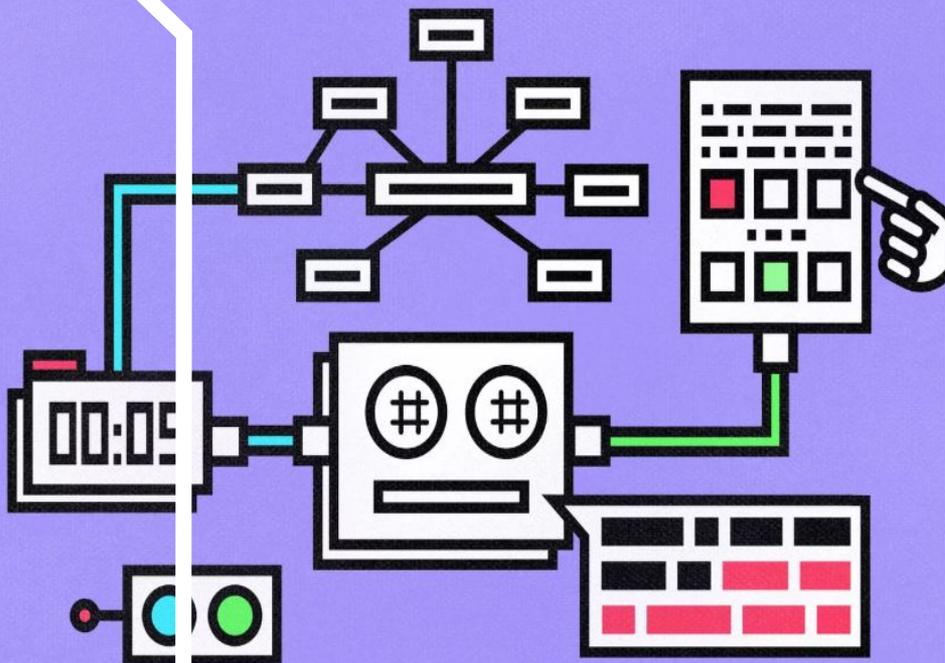
https://philosophy.hku.hk

PRACTICALLY IMPOSSIBLE

Others have argued (e.g. Judea Pearl) that implementing a Chinese Room in practice would be ridiculously hard -- the rule book would be too long. In a finite world, you'd need a **model of language.**
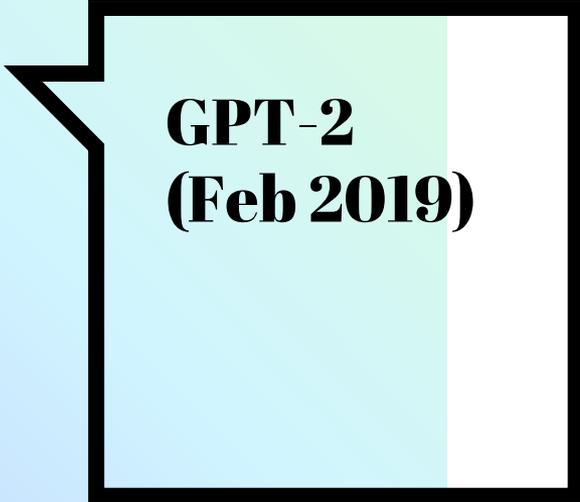
Fast Forward to February 2019...

THEVERGE

REPORT

OPENAI'S NEW MULTITALENTED AI WRITES, TRANSLATES, AND SLANDERS

A step forward in AI text-generation that also spells trouble

By James Vincent | Feb 14, 2019, 12:00pm EST

GPT-2 (Feb 2019)

The GPT-2 language model by OpenAI generates *believable*, lengthy texts from a single prompt -- essays.

- Creators choose *not to release* the full model *for fear that it will be misused.*
- Among the concerns: Could be used to create numerous "sock puppet" accounts, where you think there's a human being generating content

Language Models & Education:

Fake essays/ homework

https://twitter.com/Miles_Brundage/status/1097369503736852480

"DeepFakes" (early 2018)

Jordan Peele controlling Obama:

also search on: "Everybody Dance Now"

StyleGAN (Dec 2018)

Dec 2018: NVIDIA researchers release StyleGAN, which can synthesize believable images of...

- http://thispersondoesnotexist.com

  (or http://www.whichfaceisreal.com)

- https://thisrentaldoesnotexist.com

GANs are a true "Imitation Game"...

The Imitation GAN

D: Detective

Generative **A**dversarial **N**etworks
(Goodfellow, 2014)

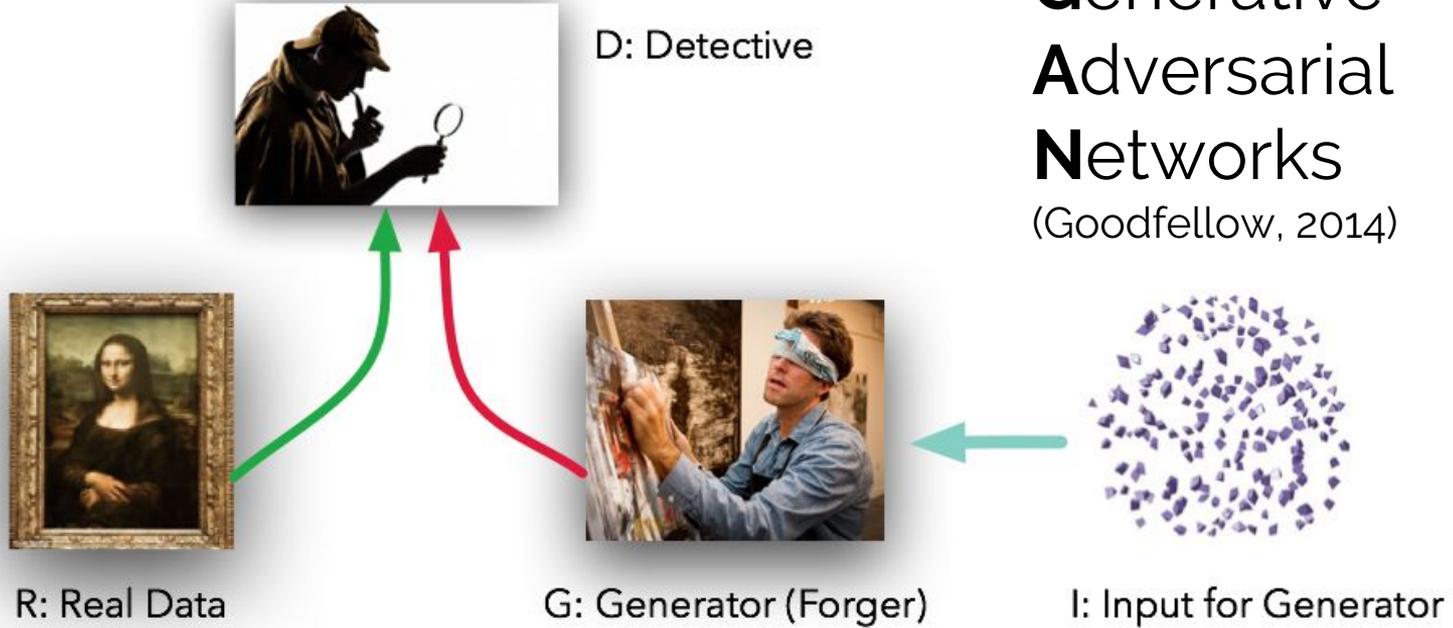R: Real Data

G: Generator (Forger)

I: Input for Generator

Image credit: Dev Nag

So…

- **Synthesized profile picture**
- **Synthesized text content**
- **Synthesized video**

==> Fake 'people' online.

- **What would that be like?**
- **What are the risks?**
- <u>**How would you verify who's human & who's a bot?**</u>

…and now we're back to Turing's question.   Enjoy the film!