GPT-3 can replicate same biases that a person have. You can simulate people in GPT-3 and they might respond with uncanny simlairty to real people in real life

Bias: racial bias. Immediate perceoption that we don’t really find.

2 approach to bias

* do things right: we need to be aware of bias and correct them. To be right and correct
* for descirptioving population. Suppose you want to predict how a population will responds to a situation. You must take in account bias to have a good prediction

Being bias in certain way is human. Complex patterns are display along different way of the same population.

GPT-3 train with demografically specific data.

Microsoft chief scientific officer, Eric Horvitz about **deepfakes** two main risks

**The US govermeent comes up with an AI “Bill of Rights” (minimum enforcement)**

Approach different to EU

EU discussing to have a proper act that will be a law in every country in EU.

Right now in USA is as guide-lines/suggestions that company should follow. USA approach main concerne is to have a moreal situation but do not interfere with business

# Computing Machinery and Intelligence

Traditional question: *can machines think?* Turing replace the question with another one that he think is important. Replace because is a very difficult question and the terms are too vague (missing definition of *think* that we could all agree at). Turing aims to separate appearance form intelligence and separate intelligence from the abstract and the practical demonstration

Ultimately the question is too vague and meaning-less. Turing dismiss a lot of traditional tought. ⟹ The imitation game that is a thought experiment.

* A: man ⟹ try to get C to make mistakes (lies)
* B: woman ⟹ try to help C
* C: interrogator ⟹ try to understand who is who

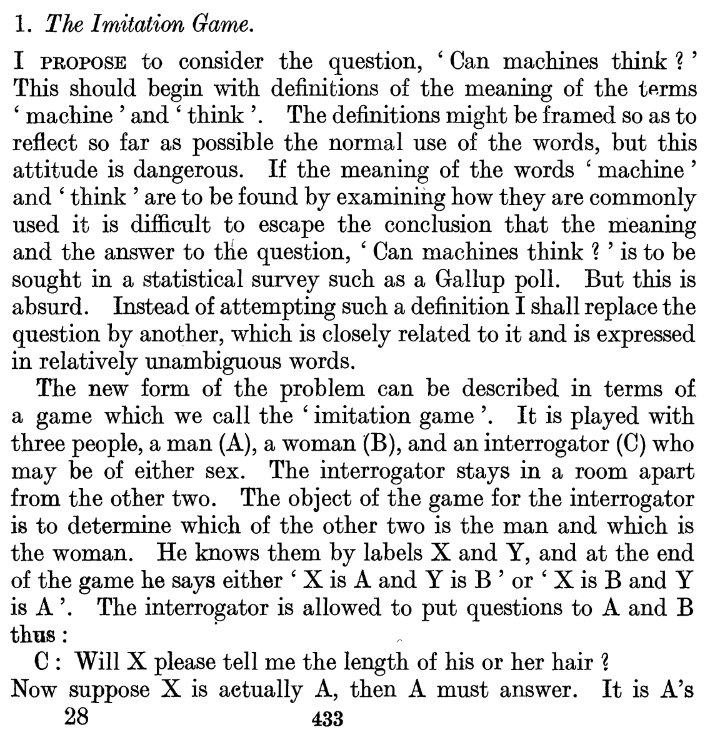
C don’t know who is the male and who is the female, C knows A and B as X(A) and Y(B). The interrogator needs to figure out who is who

Suppose you substitute A with a machine. Will C be wrong the same number of times as before (with no machine)? *Will C’s answer be better or worse when A is a machine vs when A is a human?*

[Turing - Computing Machinery and Intelligence.pdf](https://drive.google.com/file/d/1Z7aCyM8vA0kGOsPgNtomKieZIQW057tN/view?usp=sharing)

Alan Turing: decifres Enigma (WWII), create notion of Turin’s machine. Use to prove some limititative results in computer science.

Mathematician both theoretical and applied

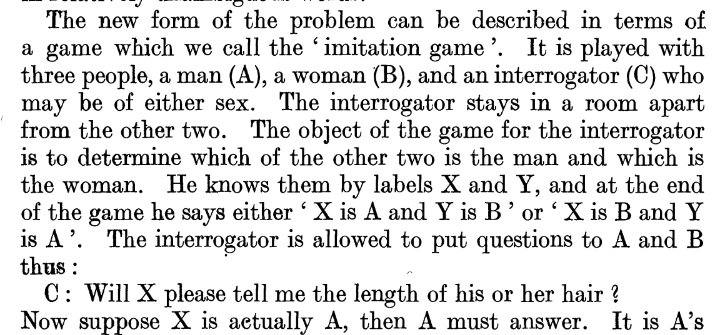


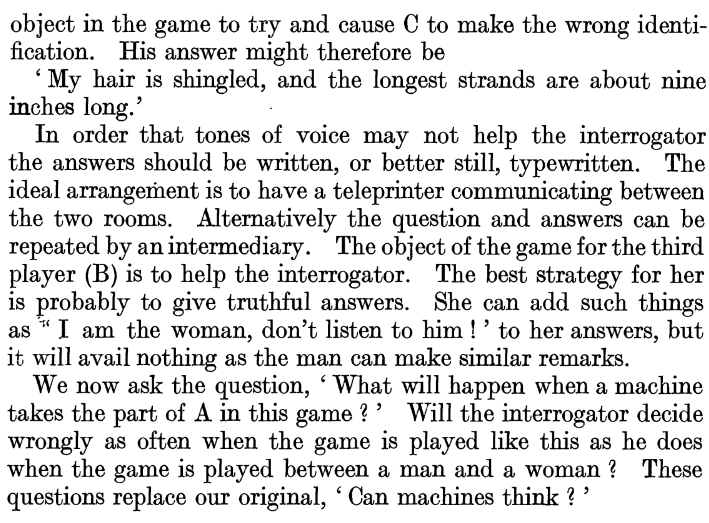
*Can machine think?* seems to equate intelligence with thinking.

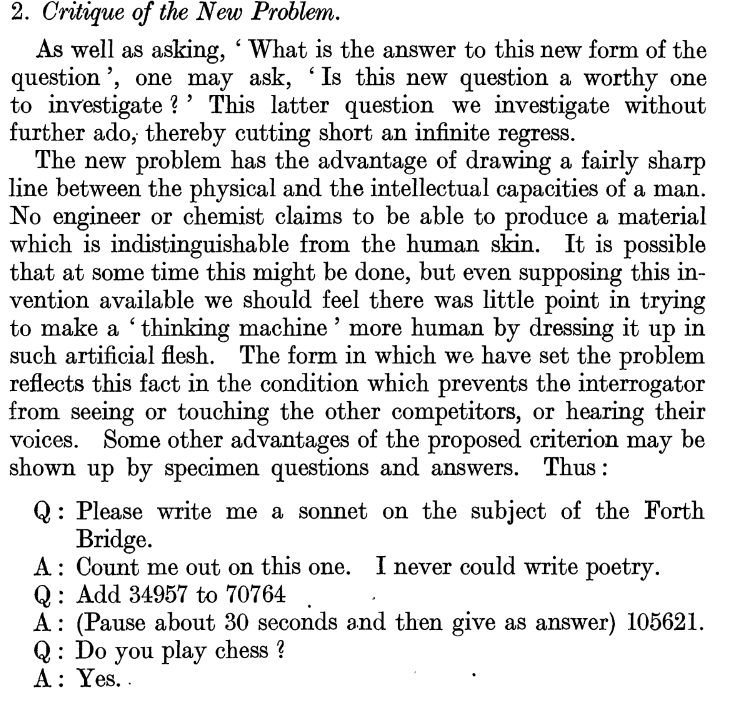
*Imitation game*. The question before is too big, hard to answer because terms could be ambiguous. Substitute the theoretical question with something else that is closely related and is unambiguous. Closely related? not clear how, there is a gab between. The things are not follow

Closely related ≠ equivalent

Conditioning: bell - food for dogs. Behaviouralism: human mind is a black box and we can intervene on it with exteranl stimuli or by study it by considering the outputs

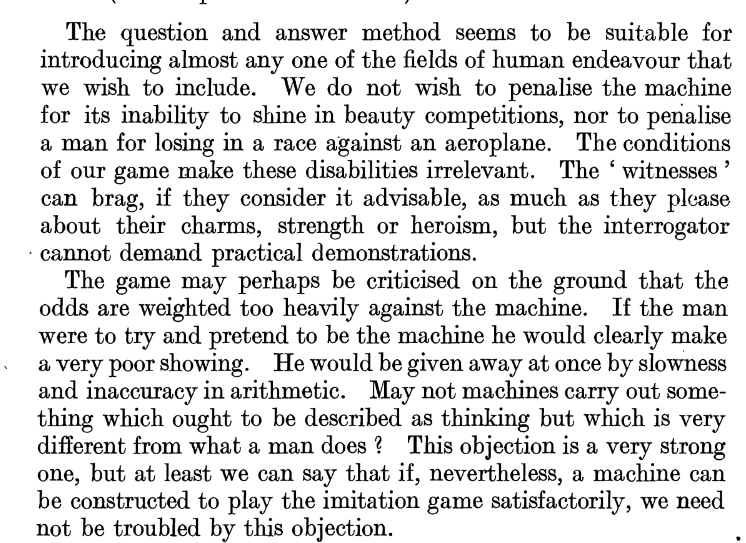




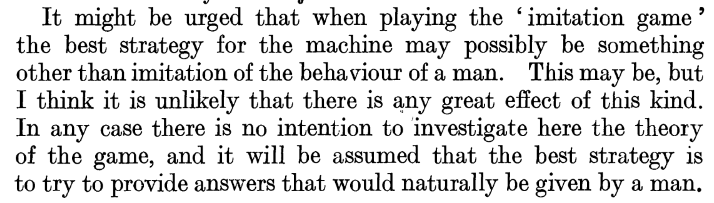


Starts enumerate some advantages and disadvantages.

To investigate something if you question an assumption and what it depends on, and what it depends on that depends on… can not going backward forever



The interrogator can no ask practical demonstration. The man is try to help the interrogator and the machine try to fool the interrogator. The text seems to bias against the machine.



Maybe the best strategy for the machine is to fool the interrogator to consider the human. Maybe main strategy would be to imitate not a man but still try to fool the human.