From human intelligence to artificial smartness

The undone science of artificial intelligence

Alex Kabbach[1]

Undone Computer Science – Nantes

07.02.2024

[1]PhD candidate in linguistics, University of Geneva

My argument, in brief

What is artificial intelligence?

VOL. LIX. NO. 236.] [October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as

Back to (Turing, 1950)

What is the Imitation Game a.k.a the Turing test?

An **indistinguishibility** test between humans and machines, consisting in:

- a **human** evaluator sitting alone in a room, communicating with
- two participants (one **human**, one **machine**) through a text-only channel, having to decide
- which one is the **human**, which one is the **machine**

The machine **passes** the test if:

- on a sufficient number of occasions, an average human judge cannot correctly identify which one is which among the two participants

The goal of artificial intelligence

Turing (1950) sets the goal of the field of AI straight: the purpose is to model **human behavior** so as to pass the Turing test

How to interpret "intelligence" is Turing's conception of "machine/artificial intelligence":

- an ability to "think" that one possesses by virtue of being human; so that
- **intelligent** behavior = **human** behavior

The question

Is it enough for machines today to display

intelligent/human behavior?

The answer

Is it enough for machines today to display
intelligent/human behavior?

NO

Machines need to be **smart**

What is smartness?

Smartness: a *normative ideal* of human behavior; how people **ought** to behave rather than how they **do** behave in practice.

Example of "smart behavior": **perfect** spelling

Intelligence vs. smartness

Contrast:

- **Smart behavior**: perfect spelling
- **Intelligent/human behavior**: imperfect spelling

The problem

People make mistakes – that's what makes them human

Consequence:

- To pass the Turing test, we need machines that make mistakes
- Who needs machines that make mistakes today?

The claim

Artificial smartness has replaced artificial intelligence

Overall

Overall

1. The **paradigm**: the “correctness principle”;
2. The **context**: normalism and the history of statistics;
3. The **consequences**: ChatGPT and the Turing test;
4. The **reasons**: why artificial smartness?
5. The **perspectives**: what doing artificial intelligence means.

The paradigm

The correctness principle

My own opinion about the aim of AI started from the vague feeling that traditional computational systems are based on a design principle that makes them very different from the human mind, and that this principle can explain many other differences between the machine and the mind: **A program is traditionally designed to do something in a predetermined correct way, while the mind is constructed to do its best using whatever it has**.

Wang (2019, p.16)

On Defining Artificial Intelligence

The correctness principle

A machine is traditionally designed to do something in a predetermined correct way

The correctness principle: consequences

Two consequences:

1. Every task is framed as a **puzzle** with a ("correct") **solution**, rather than a **problem/situation** with a potential **response/behavior**;
2. Machine behavior is framed in "normative" terms, as being either **correct** or **incorrect** with respect to the task at hand.

We touched upon that during the conference:

- Marie-Claude and Derek: how we need (and what it means) to make a program "correct/perfect";
- Pierre: an artifact is always designed for a **purpose**.

The correctness principle: what is "correct"?

Intelligence and **smartness** correspond to two different ways of characterizing what is "correct", associated with two different "normative ideals" of human behavior:

1. **Intelligence** = a normative ideal of **normality**, where "correct" machine behavior means *normal/typical* human behavior (e.g. imperfect spelling);
2. **Smartness** = a normative ideal of **exceptionality**, where "correct" machine behavior means *exceptional* human behavior (e.g. perfect spelling)

The context

The statistical origins of AI

AI is statistics on steroids.

Broussard (2018)

Artificial Unintelligence

Statistics: a paradigm change in the philosophy of Science

Normality displaced the Enlightenment idea of human nature as a central organizing concept

[. . .]

The idea of human nature was displaced by a model of normal people with laws of dispersion.

Hacking (1990)

The Taming of Chance

Normality as a central organizing concept

Normality is [...] *both timeless and dated, an idea that in some sense has been with us always, but which can in a moment adopt a completely new form of life.* [...] *As a word, 'normal'* [...] *acquired its present most common meaning only in the 1820s.* [...] *Its opposite was the pathological and for a short time its domain was chiefly medical. Then it moved into the sphere of—almost everything. People, behaviour, states of affairs, diplomatic relations, molecules: all these may be normal or abnormal. The word became indispensable because it created a way to be 'objective' about human beings.*

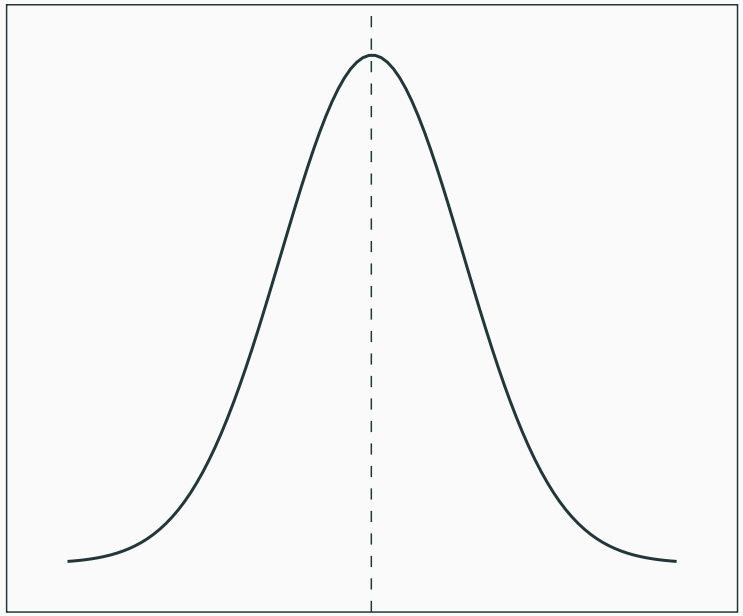
Hacking (1990)

The Taming of Chance

The statistical interpretation of normality

The "normal" as:

- the *standard*;
- the *typical*;
- the *frequent*;
- the *usual*;
- the *common*;
- the *average*.



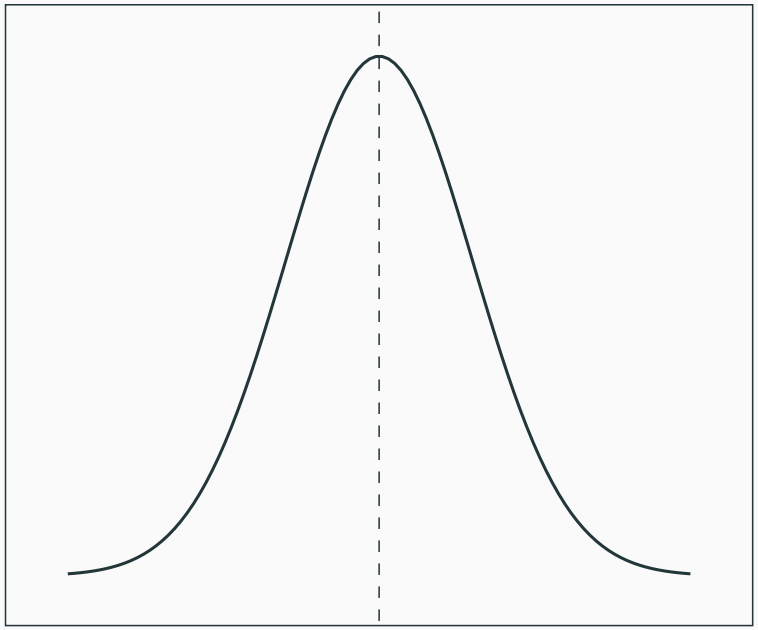
The normal curve

Then: two different normative conceptions of normality

Quetelet and the normative ideal of normality

When the normal/average is the **correct**, the **right**, the **good**

Adolphe Quetelet (1796–1874)



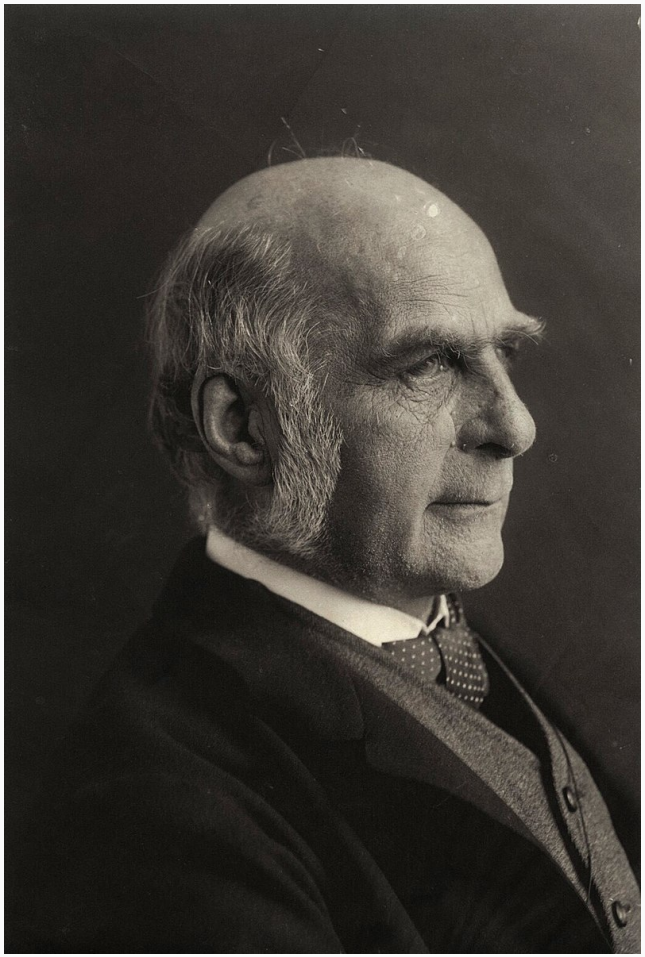
The "Quetelet Index"

a.k.a

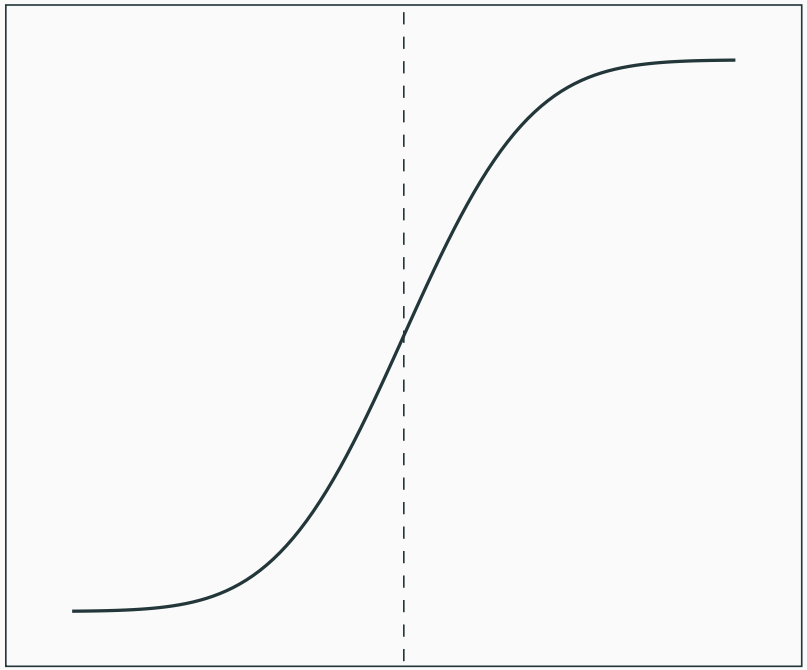
"BMI" (Body Mass Index)

Galton and the normative ideal of exceptionality

When the normal/average is only the **mediocre**



Francis Galton (1822–1911)

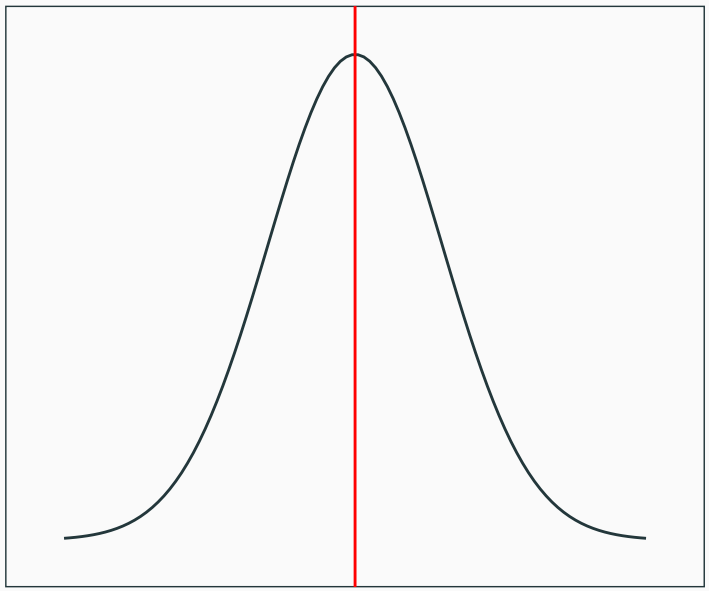


The cumulative normal curve

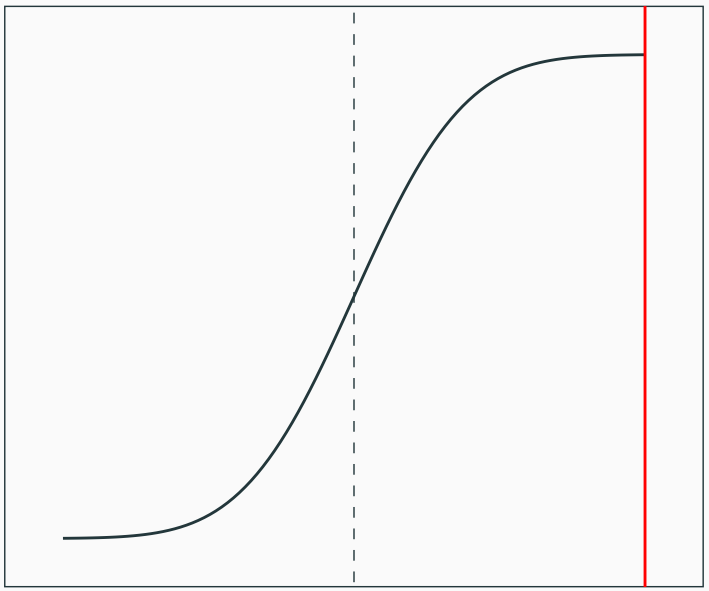
Intelligence vs. smartness / normality vs. exceptionality

Artificial **intelligence** or artificial **smartness**?

Two different normative ideals of human behavior



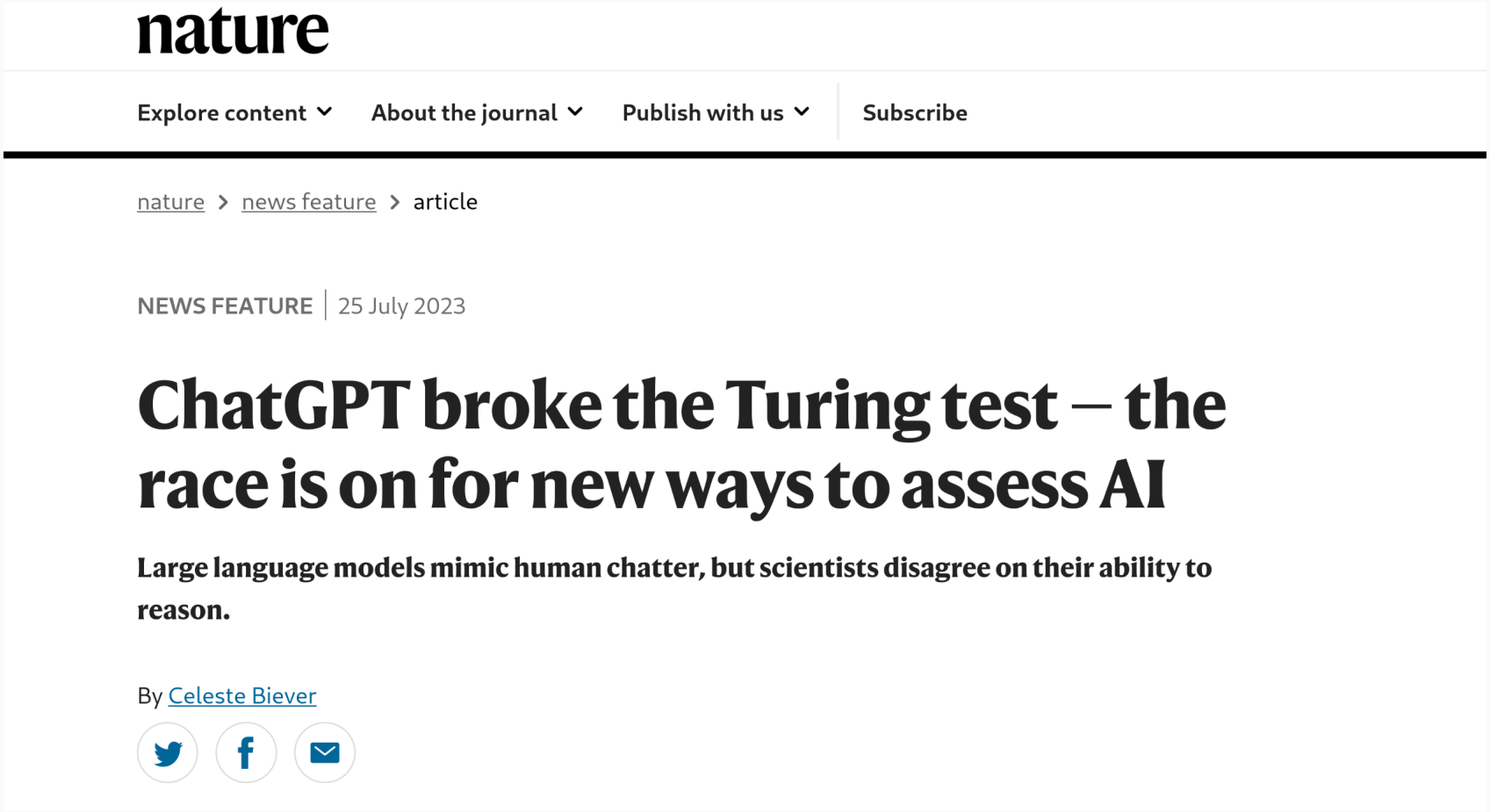
intelligence/normality



smartness/exceptionality

The consequences

Did ChatGPT break the Turing test?



nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature › news feature › article

NEWS FEATURE | 25 July 2023

ChatGPT broke the Turing test — the race is on for new ways to assess AI

Large language models mimic human chatter, but scientists disagree on their ability to reason.

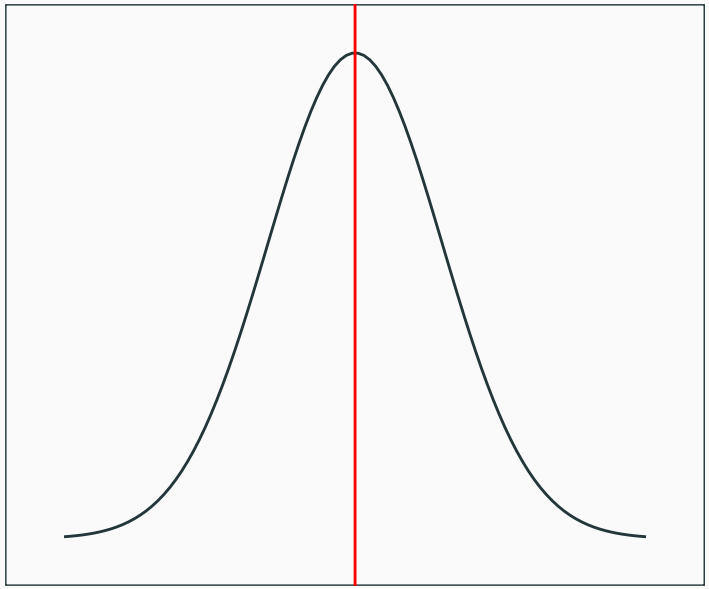
By Celeste Biever

(Biever, 2023)

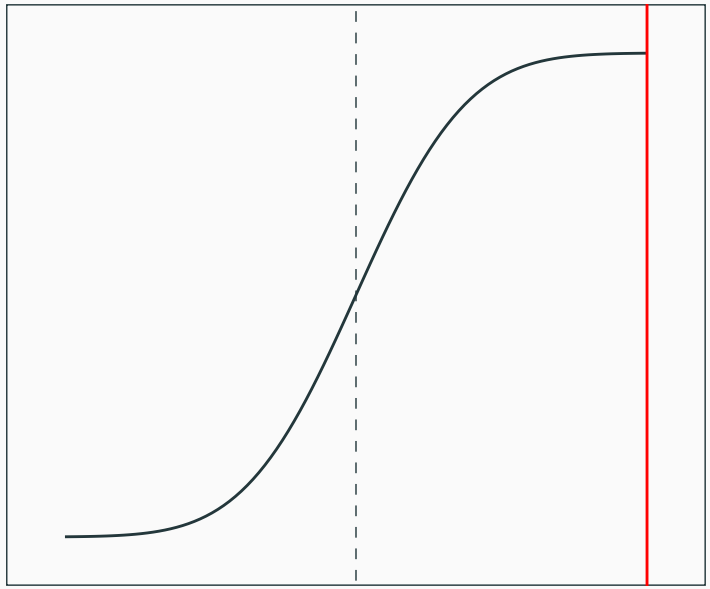
Intelligence vs. smartness / normality vs. exceptionality

Artificial **intelligence** or artificial **smartness**?

Two different normative ideals of human behavior



intelligence/normality



smartness/exceptionality

Two different and irreconcilable scientific projects

My reasons for being skeptical

ChatGPT does not make spelling mistakes

> *It is interesting here that the judge did correctly identify the human entity as there were a lot of spelling mistakes in their discourse and the conversation was quite stilted.*
>
> [. . .]
>
> *the occasional spelling mistake seems to add human credibility.*
>
> *Warwick & Shah (2016)*
> *Can machines think? A report on Turing test experiments at the Royal Society*

My reasons for being skeptical

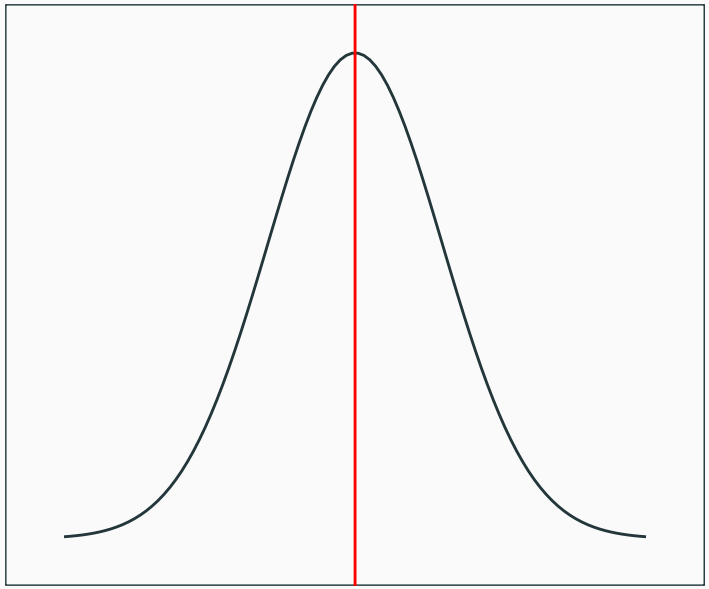
ChatGPT is too good to be human

> *Do not show that you know a lot of things—the judge may conclude that you are too clever to be human.*
>
> *Warwick & Shah (2015)*
> *Human misidentification in Turing tests*

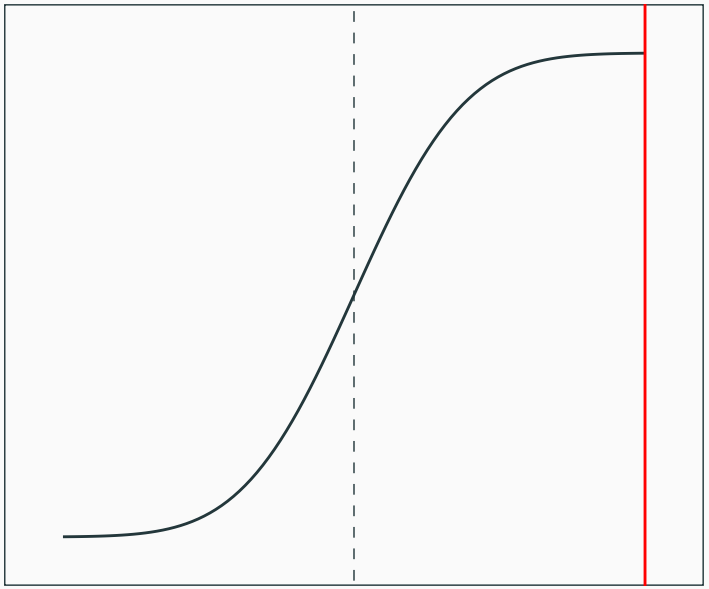
Intelligence vs. smartness / normality vs. exceptionality

Artificial **intelligence** or artificial **smartness**?

Two different normative ideals of human behavior



intelligence/normality



smartness/exceptionality

Two different and irreconcilable scientific projects

The reasons

The intuition of smartness

Historically, researchers have pursued several different versions of AI. Some have defined intelligence in terms of fidelity to human performance, while others prefer an abstract, formal definition of intelligence called **rationality***—loosely speaking, doing the "right thing."*

Russell & Norvig (2020)

Artificial Intelligence: A Modern Approach

Why smartness?

[W]e need to put aside the attempt to build a machine that can flawlessly imitate humans; for example, do we really need to build computers that make spelling mistakes or occasionally add numbers incorrectly, as in Turing's original article [...] *in order to fool people into thinking they are human?*

French (2012)

Moving beyond the Turing test

Why smartness?

Few AI researchers pay attention to the Turing test, preferring to concentrate on their systems' performance on **practical** *tasks, rather than the ability to imitate humans.*

Russell & Norvig (2020)

Artificial Intelligence: A Modern Approach

Why smartness?

[M]ost AI researchers and developers, in point of fact, are simply concerned with building **useful**, *profitable artifacts [. . .]*

Bringsjord & Govindarajulu (2022)
Artificial Intelligence – SEP

Why artificial smartness?

"Practical" ... "useful" ... for **whom**? and to do **what**?

The real question

Is there really any money/power to be gained from building machines that "make mistakes"?

The perspectives

Back to Turing: the "Arguments from Various Disabilities"

These arguments take the form, "I grant you that you can make machines do all the things you have mentioned but you will never be able to make one to do X". Numerous features X are suggested in this connexion. I offer a selection:

> *Be kind, resourceful, beautiful, friendly* [. . .] *have initiative, have a sense of humour, tell right from wrong,* **make mistakes** [. . .]

Turing (1950)
Computing machinery and intelligence

Turing's answer

The claim that "machines cannot make mistakes" seems a curious one. [...] *It is claimed that the interrogator could distinguish the machine from the man simply by setting them a number of problems in arithmetic. The machine would be unmasked because of its deadly accuracy. The reply to this is simple. The machine* [...] *would not attempt to give the right answers to the arithmetic problems. It would deliberately introduce mistakes in a manner calculated to confuse the interrogator.*

Turing (1950)

Computing machinery and intelligence

For a science of human subjectivity/variability

Does any mistake go?

- Are there "mistakes" that no human being will ever make?
- If so: how do we distinguish **human** "mistakes" from **non-human** "mistakes"?

Conclusion

Doing **artificial intelligence** requires a science of **human subjectivity/variability**, i.e. understanding how people deviate from whichever normative ideal of smartness they live by

Summary

Summary 1/2

- artificial intelligence requires modeling intelligent/**human** behavior;
- but human/intelligent behavior is not enough: machines need to be **smart**;
- the problem is: **people make mistakes**, but nobody needs machines that make mistakes;
- intelligence and smartness correspond to two different normative ideals of human behavior: Quetelet vs. Galton; **normality** vs. **exceptionality**;

Summary 2/2

- ChatGPT is not going to pass the Turing test: it is a **smart** machine, designed **not** to make mistakes;
- artificial intelligence and artificial smartness are just two different and **irreconcilable** scientific projects;
- doing artificial intelligence requires a science of human **subjectivity/variability**, i.e. an understanding of what makes a mistake "human" or not.

Questions

References

Biever, Celeste. 2023. ChatGPT broke the Turing test—the race is on for new ways to assess AI. *Nature* 619(7971). 686–689. doi:10.1038/d41586-023-02361-7. https://doi.org/10.1038/d41586-023-02361-7

References ii

Bringsjord, Selmer & Govindarajulu, Naveen Sundar. 2022. Artificial Intelligence. In Zalta, Edward N. & Nodelman, Uri (eds.), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University Fall 2022 edn. https://plato.stanford.edu/archives/fall2022/entries/artificial-intelligence/.

Broussard, Meredith. 2018. *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge, MA: MIT Press. doi:10.7551/mitpress/11022.001.0001. https://doi.org/10.7551/mitpress/11022.001.0001

References iii

French, Robert M. 2012. Moving beyond the Turing test. *Communications of the ACM* 55(12). 74–77. doi:10.1145/2380656.2380674. https://doi.org/10.1145/2380656.2380674

Hacking, Ian. 1990. *The Taming of Chance*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511819766. https://doi.org/10.1017/CBO9780511819766

Russell, Stuart J. & Norvig, Peter. 2020. *Artificial Intelligence: A Modern Approach*. Hoboken, NJ: Pearson Education, Inc. 4th edn.

References iv

Turing, Alan M. 1950. Computing machinery and intelligence. *Mind* LIX(236). 433–460. doi:10.1093/mind/LIX.236.433. https://doi.org/10.1093/mind/LIX.236.433

Wang, Pei. 2019. On Defining Artificial Intelligence. *Journal of Artificial General Intelligence* 10(2). 1–37. doi:10.2478/jagi-2019-0002. https://doi.org/10.2478/jagi-2019-0002

Warwick, Kevin & Shah, Huma. 2015. Human misidentification in Turing tests. *Journal of Experimental & Theoretical Artificial Intelligence* 27(2). 123–135. doi:10.1080/0952813X.2014.921734. https://doi.org/10.1080/0952813X.2014.921734

References v

Warwick, Kevin & Shah, Huma. 2016. Can machines think? A report on Turing test experiments at the Royal Society. *Journal of Experimental & Theoretical Artificial Intelligence* 28(6). 989–1007. doi:10.1080/0952813X.2015.1055826. https://doi.org/10.1080/0952813X.2015.1055826