Artificial Intelligence Applied to Scientific Research and Ethics: How can we leverage AI algorithms to achieve better science?

Mestrado em Energias Sustentáveis | Workshop

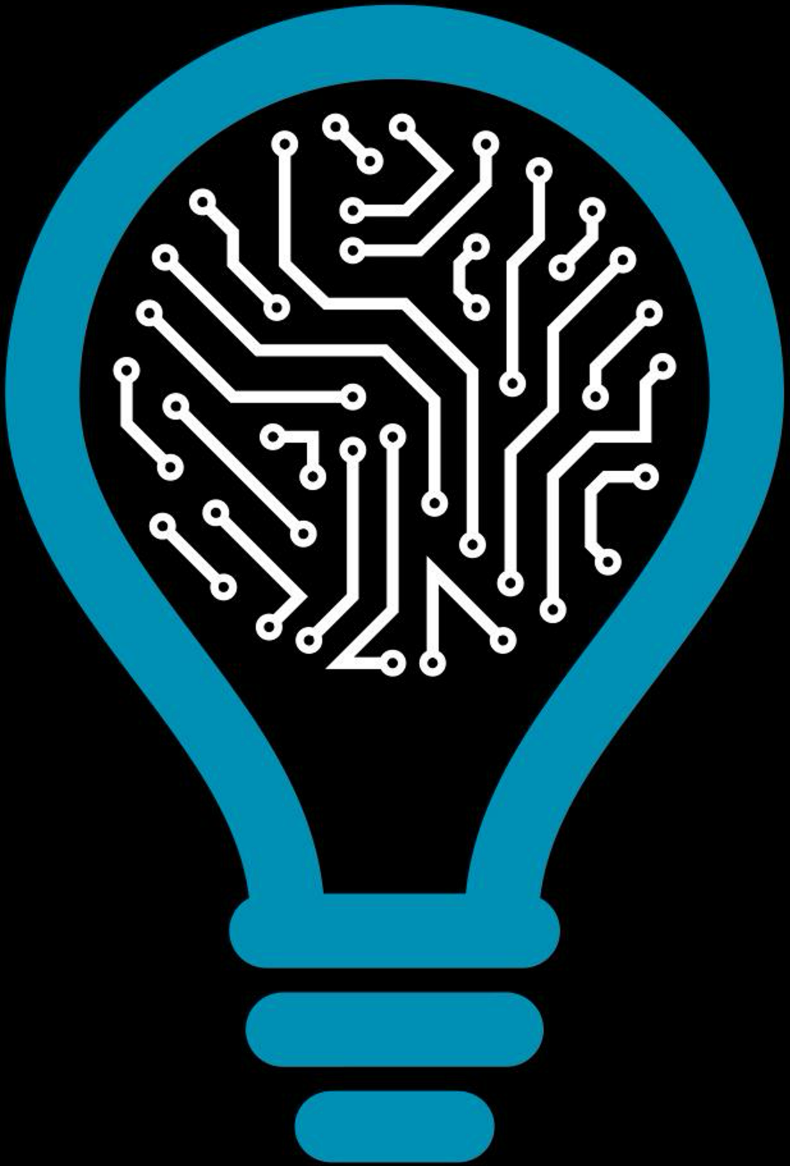
December 10, 2024 | Instituto Superior de Engenharia do Porto

Tiago Filipe Sousa Gonçalves

https://tiagofilipesousagoncalves.github.io/

INESCTEC

INSTITUTE FOR SYSTEMS AND COMPUTER ENGINEERING, TECHNOLOGY AND SCIENCE



INESCTEC

Outline

1. Modern Artificial Intelligence is Statistics on Steroids

2. Regulating Science: Research Integrity and Ethics

3. Artificial Intelligence meets Science

4. Surfing the Waves of AI: Take-home Messages

1. Modern Artificial Intelligence is Statistics on Steroids

In the beginning, we had rules

- In the beginning, **artificial intelligence systems were based in algorithms**:
  - An algorithm is a **set of instructions** that the system will follow to **achieve a certain goal** (direct programming)[1]
  - These **explicit** rules were often based on **domain knowledge**
  - Hence, they were "easy" to **explain** and to **understand**

- Nowadays, we use the available data to automatically learn **programs/functions**:
  - In machine learning, we **learn from data and make predictions** (indirect programming)[1]
  - These algorithms work by **optimising an objective function**
  - Hence, the "rules" often are **implicit** and **difficult to understand**

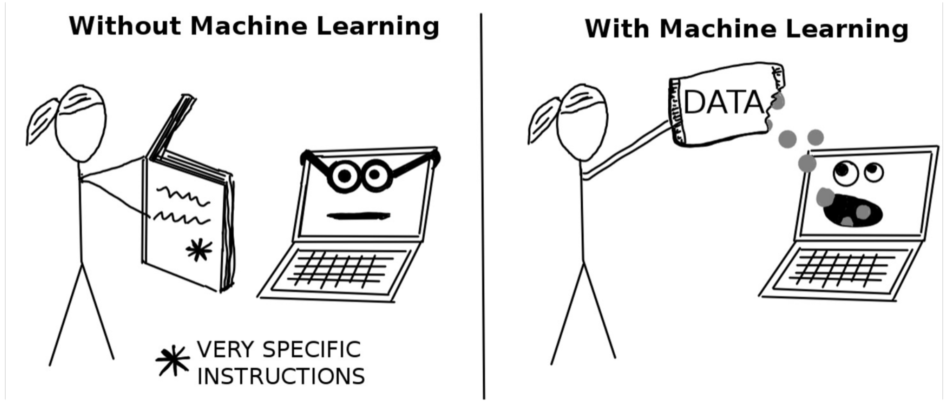


Figure - Algorithms vs Machine Learning (Image from [1])

Sources: [1] Christoph Molnar "Interpretable Machine Learning A Guide for Making Black Box Models Explainable"

Today, we have data[1]

- **The paradigm is changing:** most of the daily tasks and services can now be performed with the aid of **digital applications** or **gadgets**

- High-tech companies such as Google, Facebook, Netflix or Amazon **have access to huge amounts of data from several data sources and users**:

  - This phenomenon suggests that the *business of data* will become a **significant sector of the global economy**[2]

  - There are several **open-source data sets with millions of entries** (e.g., ImageNet[3])

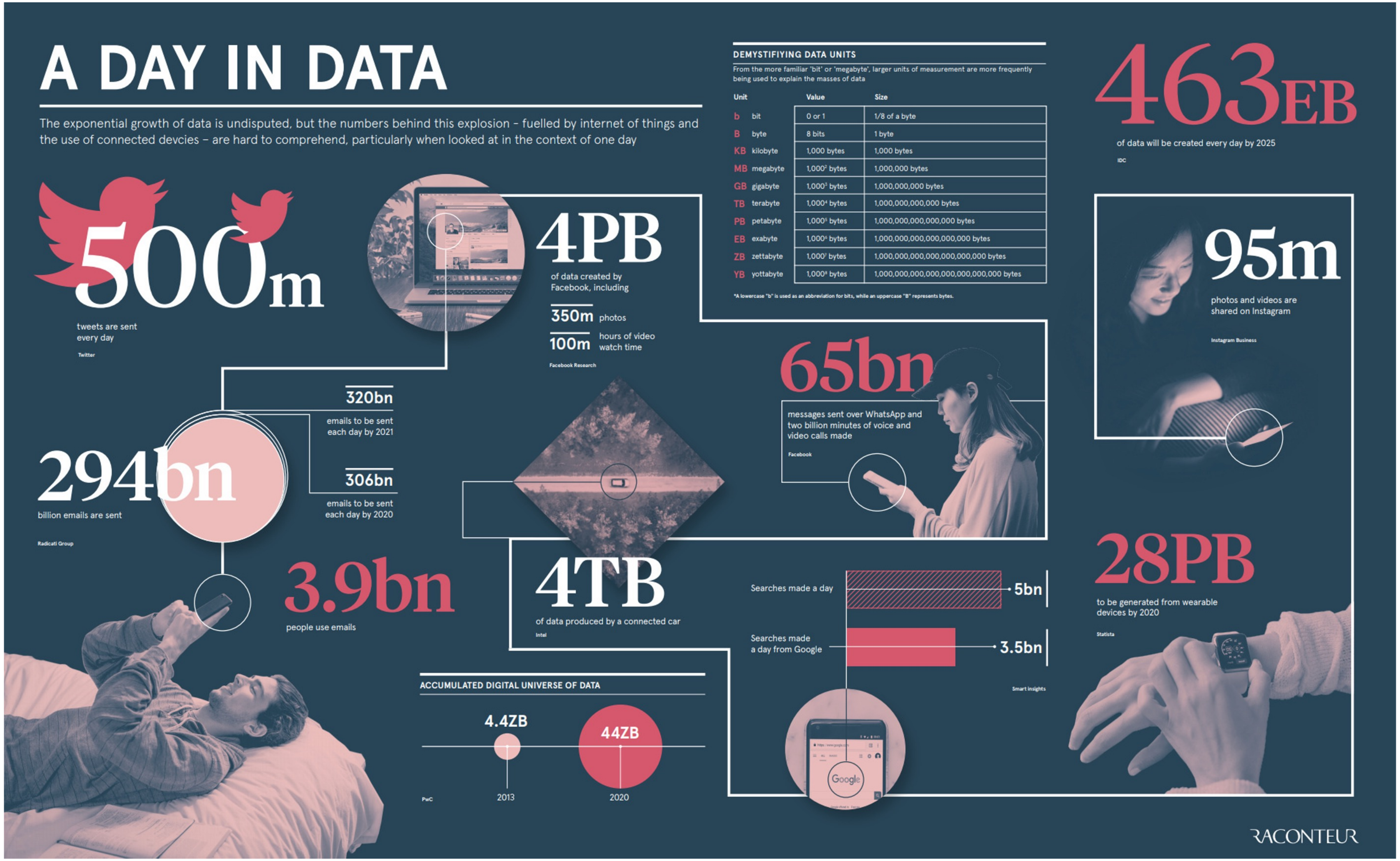
- Data is referred as **the new oil**[4]

  - The main impact on humanity is related **to the way data can improve our lives**

  - **A proper management process of the "dark side" of data must be implemented**, but the **advances** in data fuels **are worth the effort**

Sources: [1] https://techjury.net/blog/big-data-statistics/#gref, **[2]** https://www.forbes.com/sites/gilpress/2020/01/06/6-predictions-about-data-in-2020-and-the-coming-decade/?sh=5cbaadf4fc36, **[3]** http://www.image-net.org, **[4]** https://www.forbes.com/sites/forbestechcouncil/2019/11/15/data-is-the-new-oil-and-thats-a-good-thing/?sh=69bb9a407304

Yes, lots of data[1, 2, 3]



Sources: [1] https://www.raconteur.net/infographics/a-day-in-data/, [2] https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/, [3] https://www.visualcapitalist.com

We have more computational power than ever

- **The fundamental concepts of artificial intelligence and deep neural networks have been around since 1940**[1]
  - Frank Rosenblatt proposed one of the first approaches to the design and training of artificial neural networks: **the Perceptron**[2]

- The development of **powerful computer processing units (CPUs)** and the leveraging of the **graphical processing units (GPUs)**[3] for computation allowed the training of deep and complex algorithms in **"human time"**

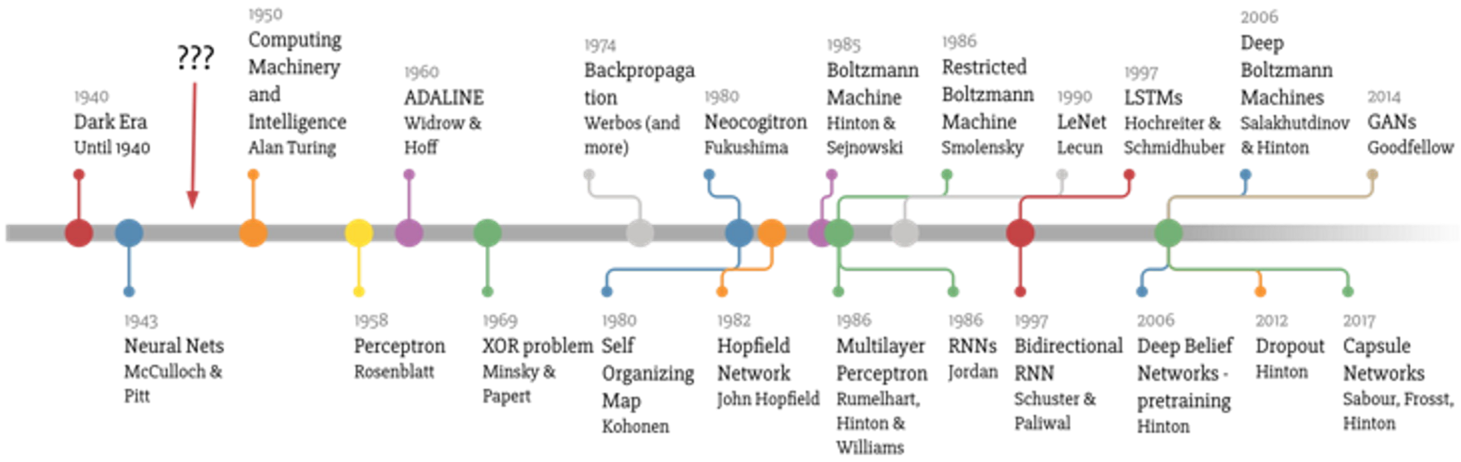


Figure - A (tentative) deep learning timeline (Image from [1])

Sources: **[1]** https://towardsdatascience.com/a-weird-introduction-to-deep-learning-7828803693b0, **[2]** Frank Rosenblatt "The perceptron: A probabilistic model for information storage and organization in the brain.", **[3]** https://www.nvidia.com/en-us/

Deep learning *versus* traditional machine learning[1]

- **Traditional machine learning** required **experts to extract meaningful features** (*i.e.*, domain-specific features) from raw data and feed them into machine learning algorithms to obtain classification/regression models:

- **Deep learning** "only" requires **raw data and labels** to achieve high-performing models, since it **automatically extracts the patterns**
  - Deep learning algorithms are suitable for **representation learning**, i.e., finding the **best representation of the data** that optimises a given optimisation objective

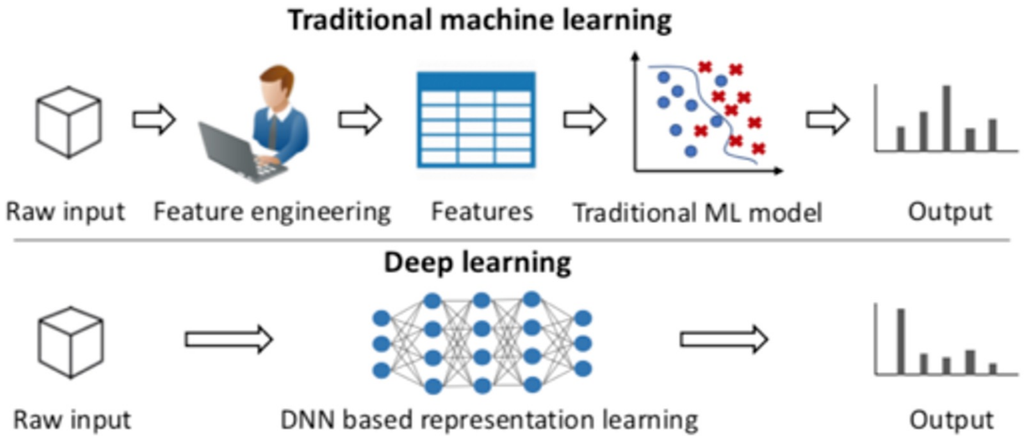


Figure - Deep learning vs traditional machine learning (Image from [2])

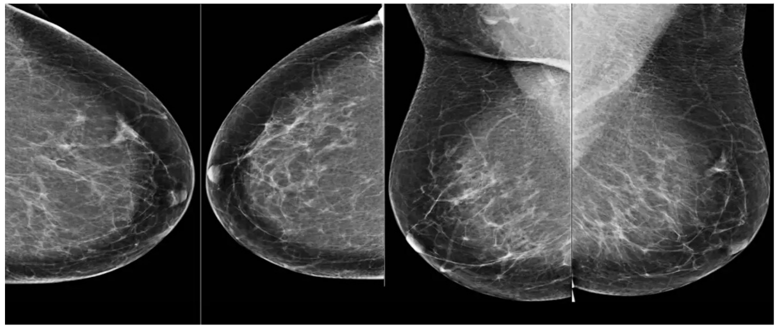
Sources: [1] Lecun et al. "Deep learning", **[2]** https://www.cc.gatech.edu/~alanwags/DLAI2016/(Gunning)%20IJCAI-16%20DLAI%20WS.pdf

Technology has been *challenging* human performance...

- There are, at least, two popular events that created a revolution in the History of AI:
    - **In 1997, IBM's Deep Blue beat the Chess World Champion Garry Kasparov**[1]
    - **In 2016, Google's DeepMind AlphaGo learn to play Go alone (i.e., through reinforcement learning policies) and beat the Go World Champion Lee Sedol**[2]

- The two events above are examples of the **(virtually) unlimited boundaries of the application of artificial intelligence** to our daily lives
    - In 2020, Google's DeepMind published a paper in ***Nature*** **suggesting that "its model was able to spot cancer in de-identified screening mammograms with fewer false positives and false negatives than experts"**[3, 4]

Figure - Medical Image Analysis: Mammograms (Image from [4])



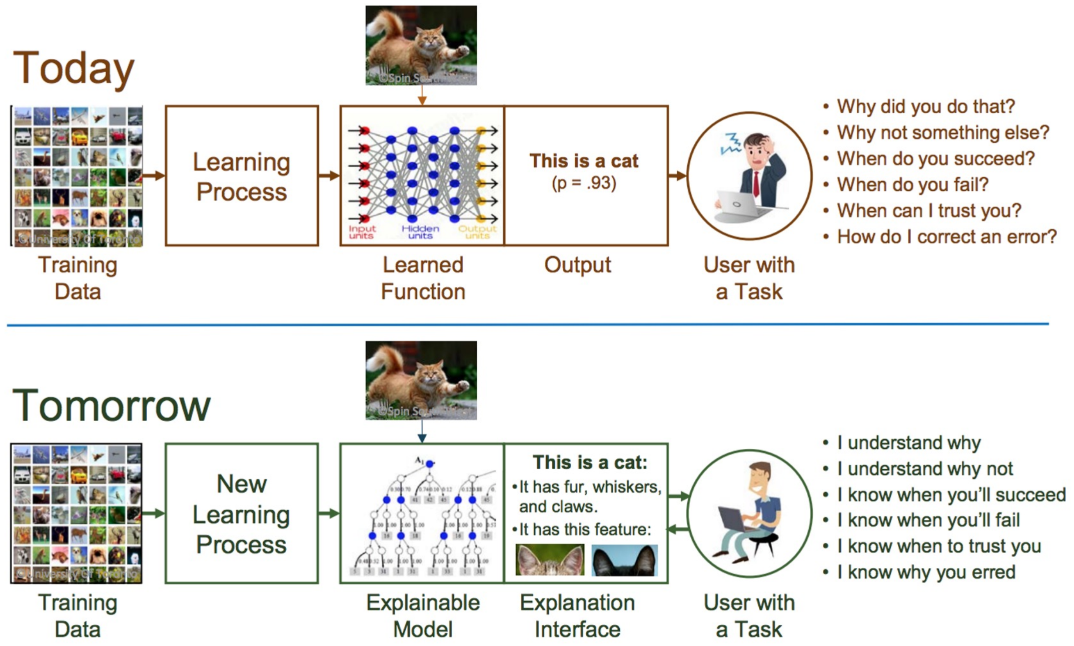
Sources: [1] https://theconversation.com/twenty-years-on-from-deep-blue-vs-kasparov-how-a-chess-match-started-the-big-data-revolution-76882, [2] https://deepmind.com/research/case-studies/alphago-the-story-so-far, [3] https://www.healthcareitnews.com/news/google-ai-platform-aids-oncologists-breast-cancer-screenings, [4] https://www.nature.com/articles/s41586-019-1799-6

Do we understand the features learned by these models?

- Even if the models achieve high performances, **it is not trivial to assure that they are learning features that are relevant for that domain** (i.e., **black box** behaviour)
  - Machine learning models **are good at extracting correlations**

- While this **may not be an issue in several domains** (e.g., recommendation systems), in others, it is of utmost importance that the **system is capable of transparently showing** the reasons behind its decisions (e.g., healthcare)

Figure - The future of machine learning algorithms
(Image from [1])



Sources: [1] https://www.cc.gatech.edu/~alanwags/DLAI2016/(Gunning)%20IJCAI-16%20DLAI%20WS.pdf

Responsible AI: Modern problems require modern solutions

- **Responsible AI** is a framework that guides how we should address the challenges around artificial intelligence from both an **ethical, technical and legal** point of view[1]
  - We must resolve ambiguity for where responsibility lies if something goes wrong!

- This framework relies on fundamental principles[2]:
  - Accountability
  - Interpretability
  - Fairness
  - Safety
  - Privacy

Figure - Responsible AI (Image from [1])

Sources: [1] https://searchenterpriseai.techtarget.com/definition/responsible-AI, [2] https://towardsdatascience.com/what-is-responsible-ai-548743369729

2. Regulating Science: Research Integrity and Ethics

My first contact with Research Integrity and Ethics happened during my PhD with a course on the topic

- I had the privilege of participating in the **3rd Edition of the Training Course in Research Ethics/ Integrity**, organised by Susana Magalhães (**Unit for Responsible Conduct in Research, i3S**)

Free Advertisement: [Training Course in Research Ethics/ Integrity - 5th Edition]

INESCTEC

This course presented 8 fundamental topics

1. Fundamentals of Bioethics and Research Ethics
2. Good Researcher what it is and why it matters
3. Preventing and managing conflicts in research
4. Authorship and Publication Ethics
5. Misconduct and Unacceptable Practices
6. Data Protection and Intellectual Property
7. Open Science
8. Science Communication and Citizen Engagement and Vulnerability and Equity in Research

Sources: xkcd "1390: Research Ethics" | Free Advertisement: Training Course in Research Ethics/ Integrity - 5th Edition

Ethics in Science: Lessons learned from World War II

1. Scientific progress does not always result in greater good for humanity (it can also lead to suffering and death)

2. Science is not a value in itself but must remain an instrument for realising human goals (the goals do not justify all means)

3. All knowledge has a practical application, so scientists must take responsibility for foreseeing the possible consequences of the knowledge they build (and preventing its harmful uses)

4. While Science answers for what we can do, Ethics will state what we should do, taking as a criterion the human value

5. Science's self-regulation is not enough to guarantee the goodness of its ends; we need Ethical scrutiny to achieve this

6. **New technologies have lost their status as inert and passive to become dynamic and active, no longer depending solely on the designs of the user**

Acknowledgements and Credits: Maria do Céu Patrão Neves

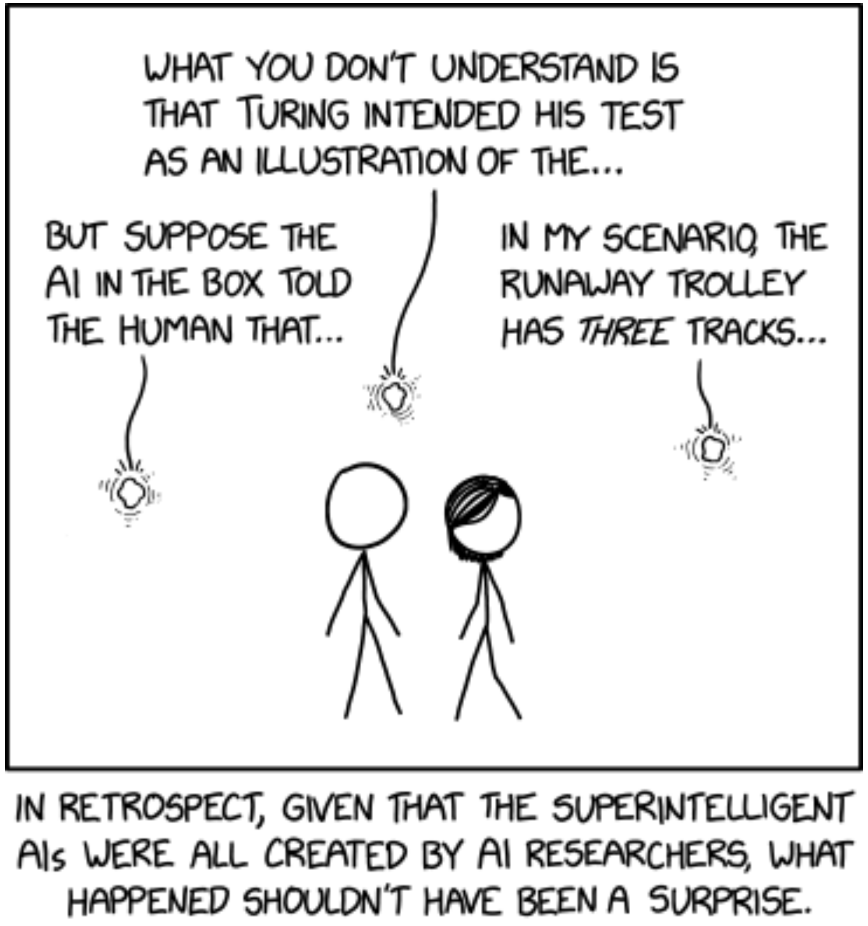
Research Integrity in Science: Why it matters

Scientific Research Integrity sets out the structuring of Ethical principles for responsible scientific research and innovation:

- Truth, rigour and objectivity
- Independence, impartiality and impartiality
- Co-operation and honesty
- Transparency and justice
- Commitment and social responsibility

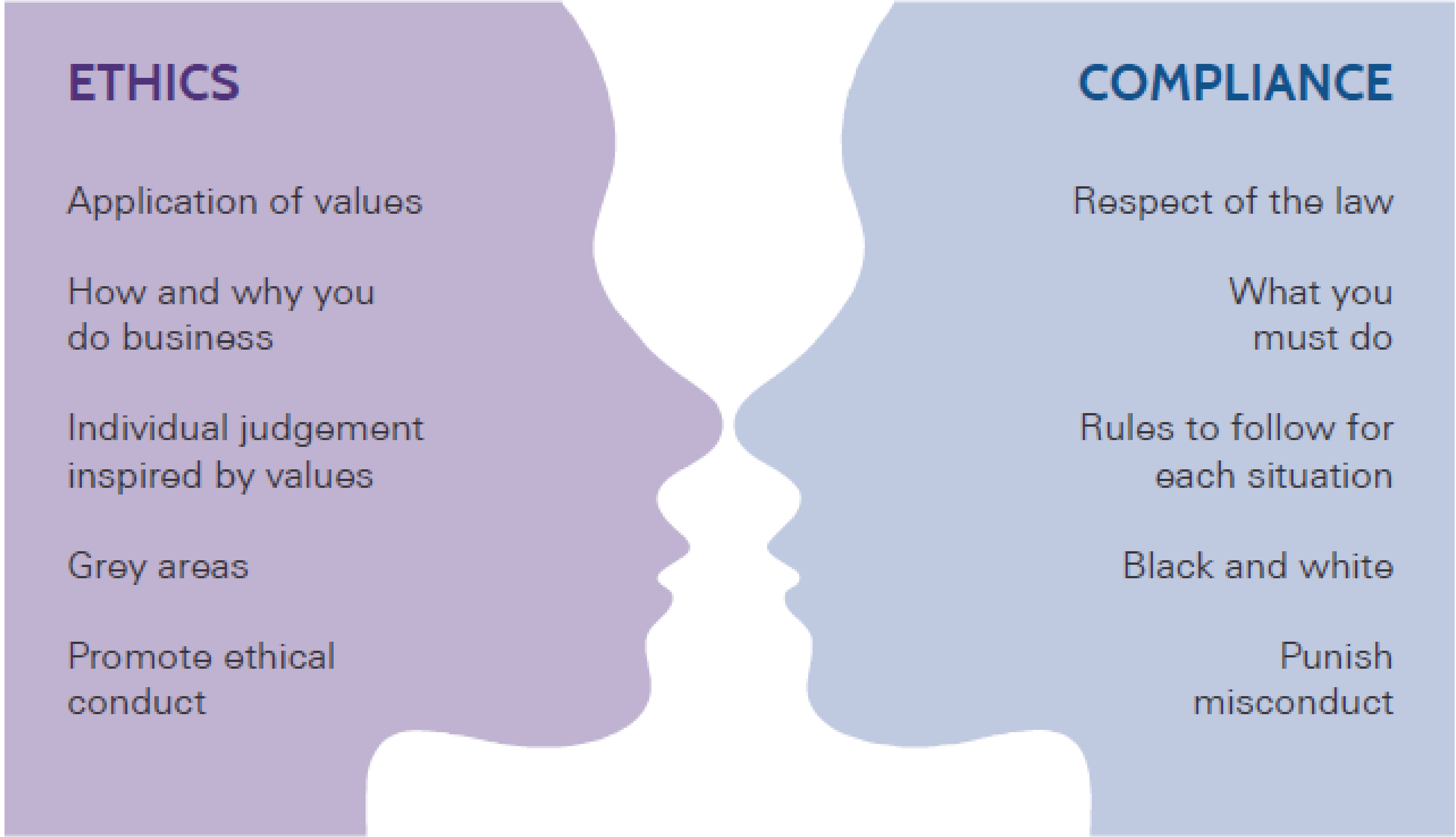
We still need an Ethical reflection applied to scientific research to:

1. Ensure that scientific progress and technological innovation conform to common morality
2. Guarantee the goodness of its aims and impacts
3. Maintain the human as the purpose of human creation



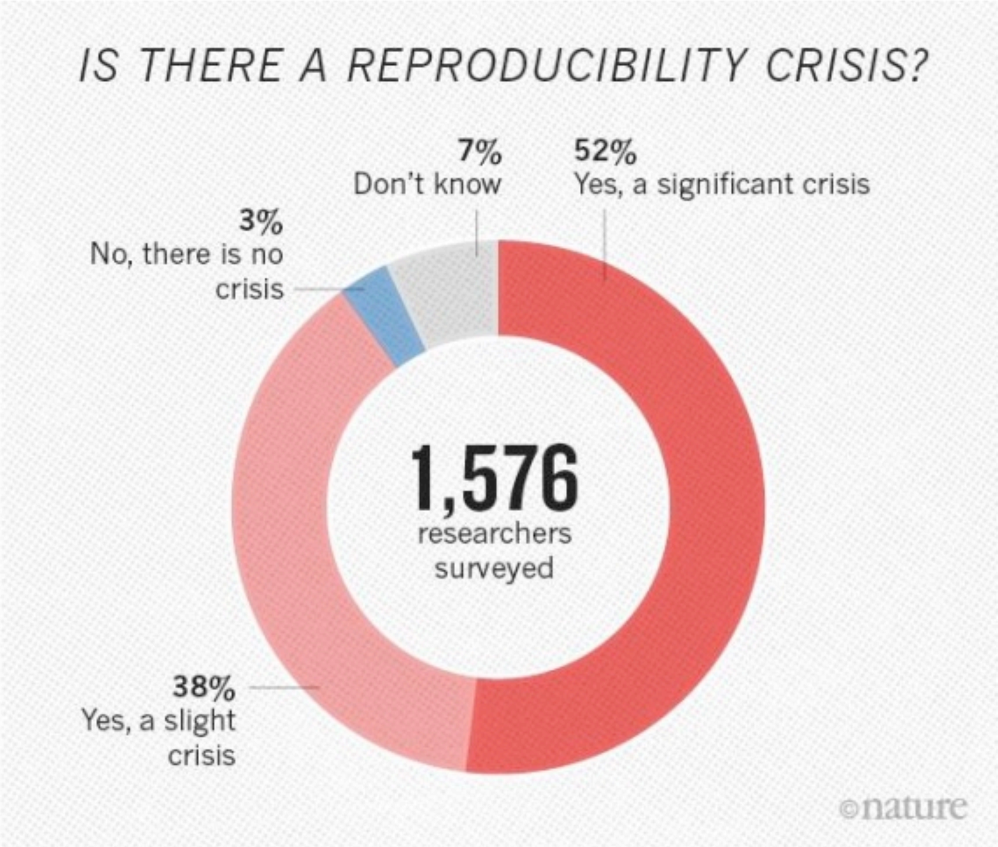
Sources: [xkcd "2635: Superintelligent AIs"] | Acknowledgements and Credits: Maria do Céu Patrão Neves

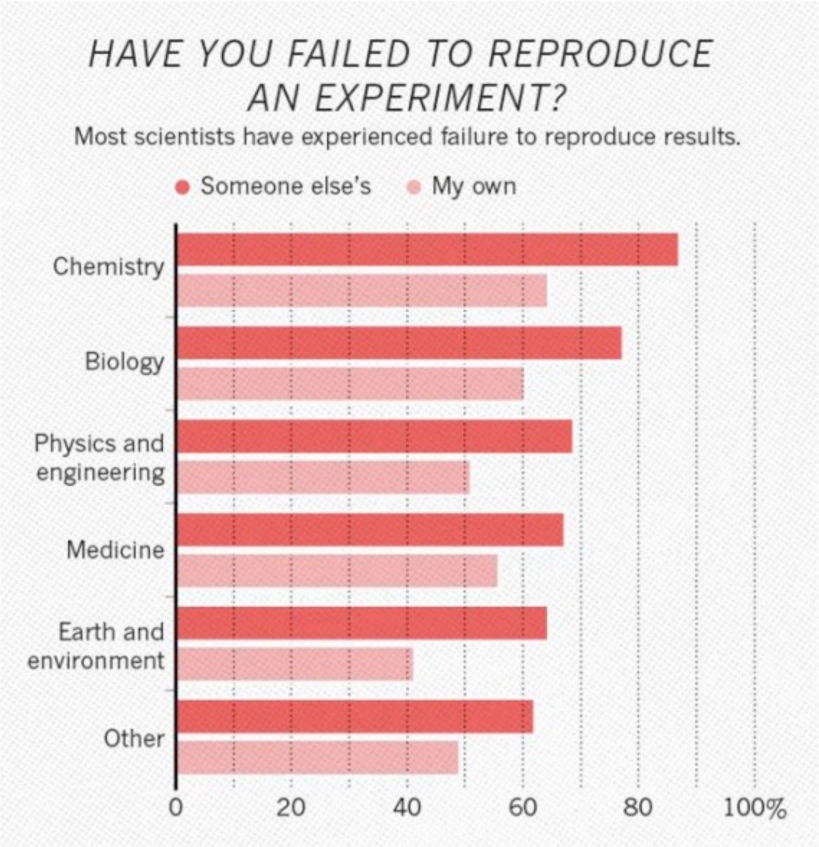
The duality between Ethics and Integrity (or Compliance)



Sources: [Ethics and Compliance Handbook 2017] | **Acknowledgements and Credits: Susana Magalhães (i3S)**

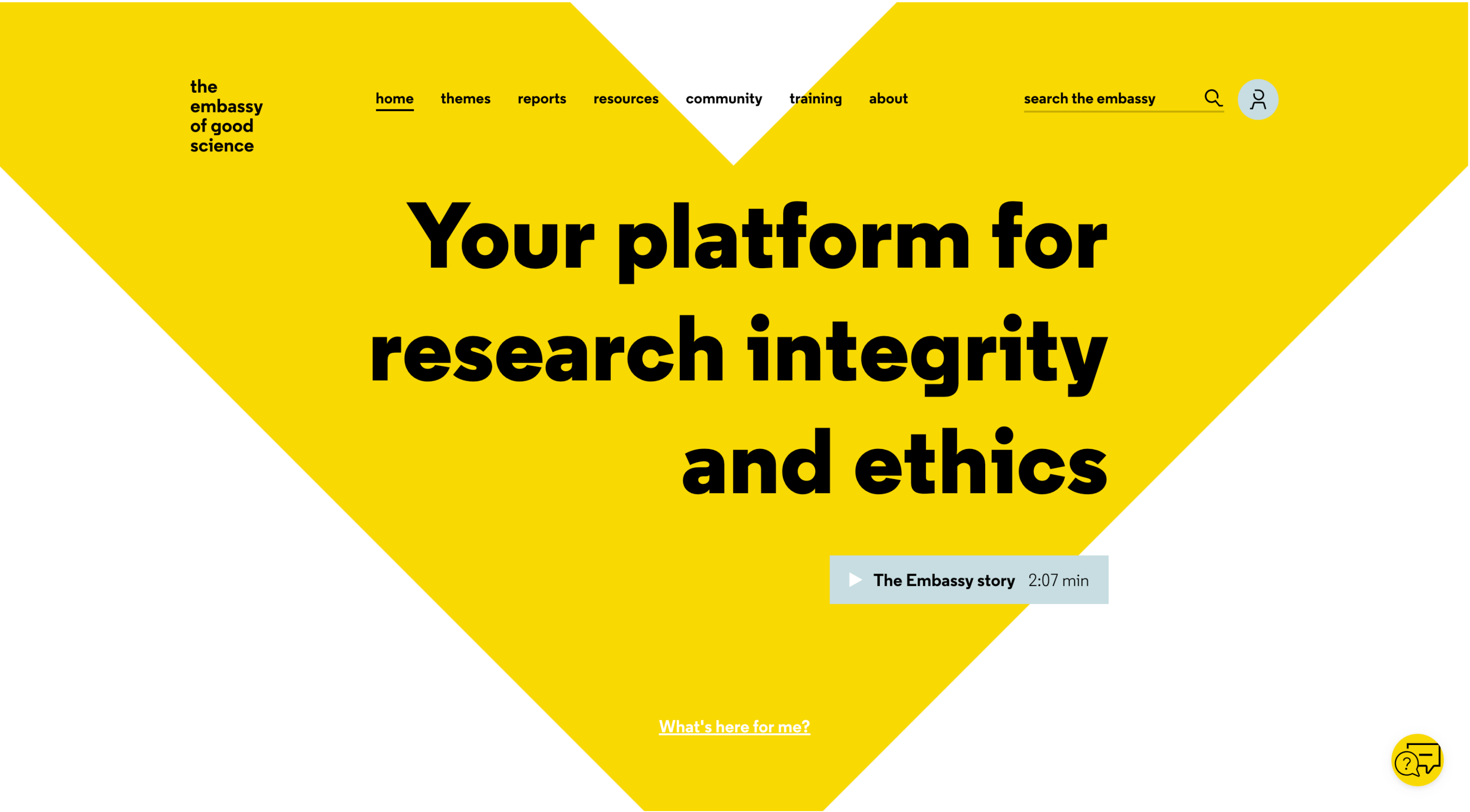
Hot Topic: Ensuring that Science is Reproducible[1]





Sources: [1] Monya Baker "1,500 scientists lift the lid on reproducibility" | Acknowledgements and Credits: Susana Magalhães (i3S)

All you need is “The Embassy of Good Science”[1]

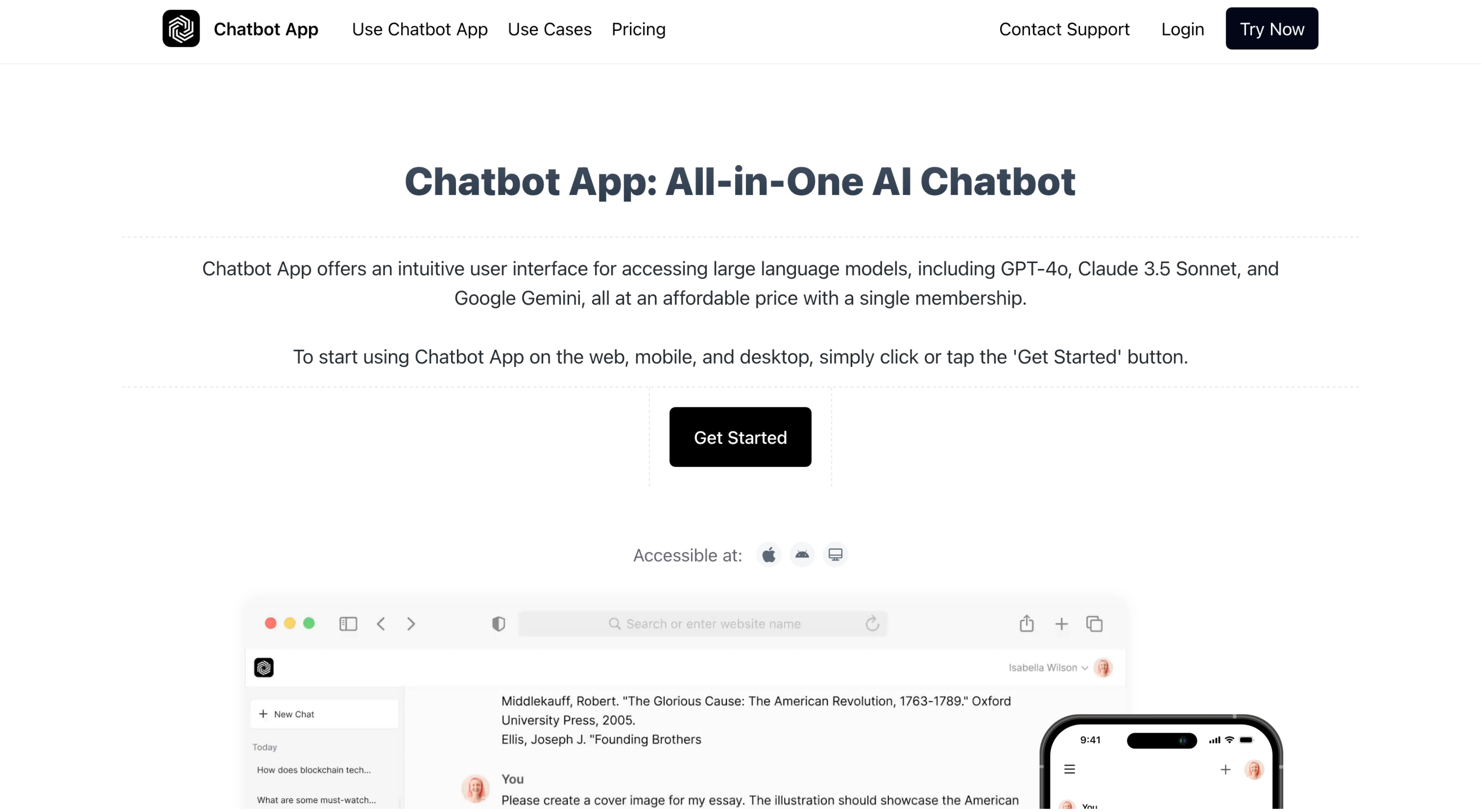


Sources: [1] The Embassy of Good Science

3. Artificial Intelligence meets Science

Should we ChatGPT all over the way?

- Nowadays, my perception is that everyone uses ChatGPT for everything, but are we using it correctly?



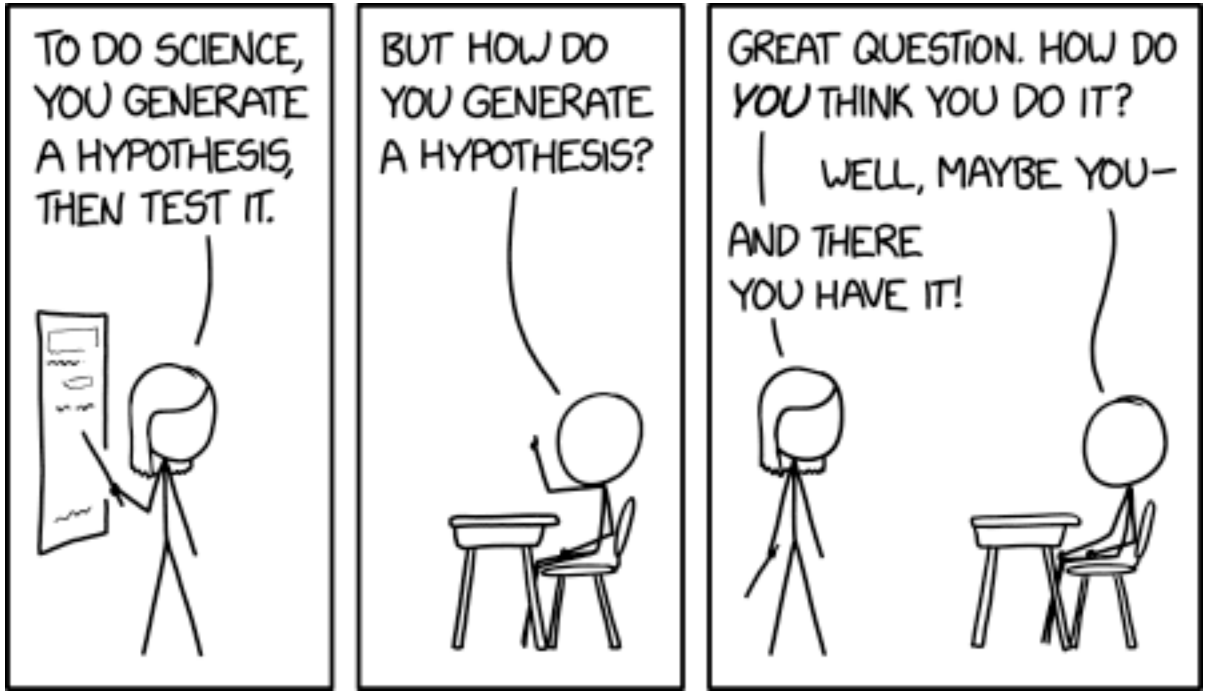
Sources: https://chatbotapp.ai/

From research to publication: the typical pipeline

Hypothesis or Idea → Literature and State of the Art Review → Experimental Protocol, Materials and Methods → Conduct Experiments, Generate and Analyse Data → Write a Report, Dissertation, Thesis or Paper → Publish and Present the Scientific Findings

Hypothesis or Idea

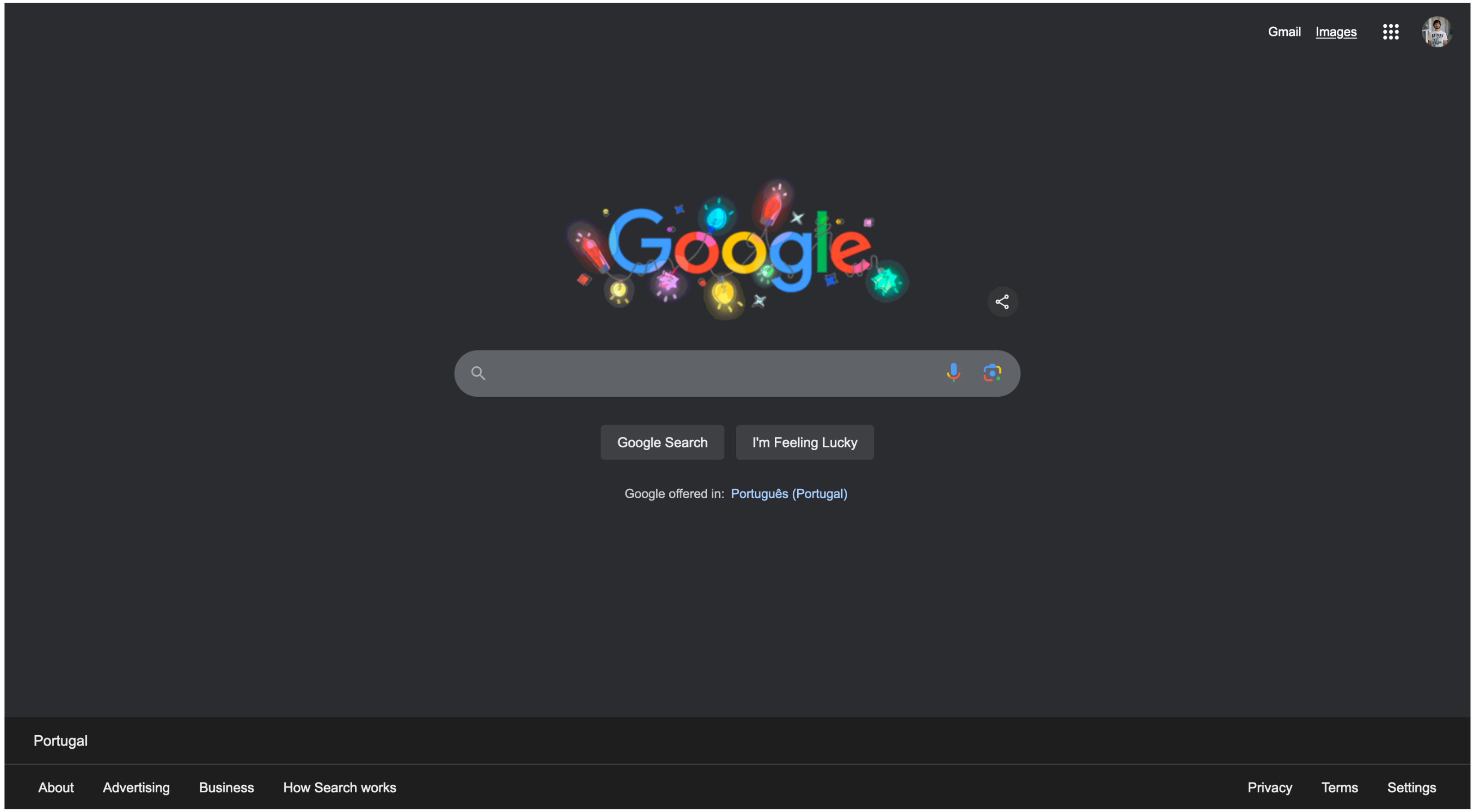
- Sometimes, you don't need AI…



Sources: [xkcd "2569: Hypothesis Generation"]

Hypothesis or Idea

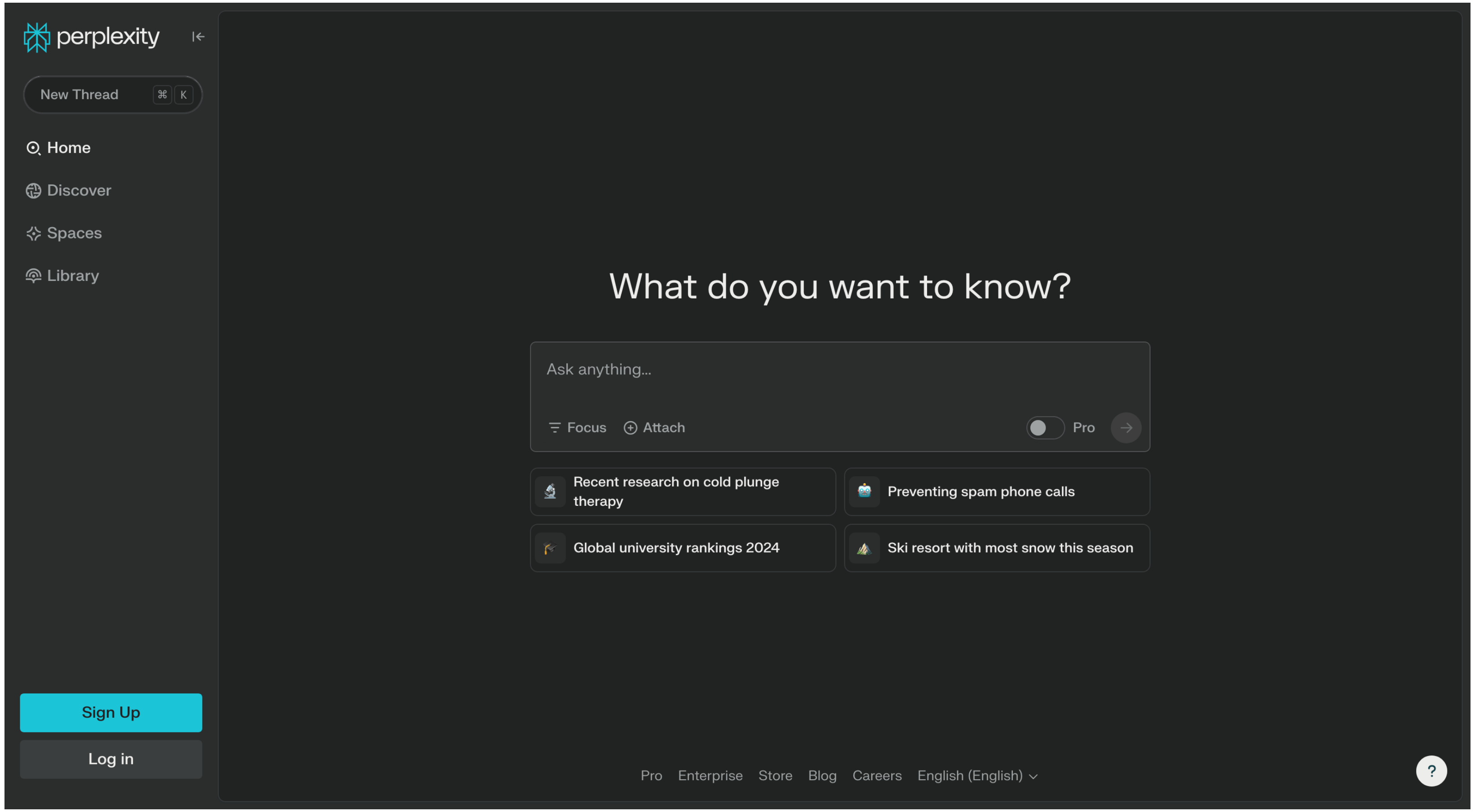
- **Google**[1] is still a nice and (fairly) good research engine



Sources: [1] https://www.google.com/

Hypothesis or Idea

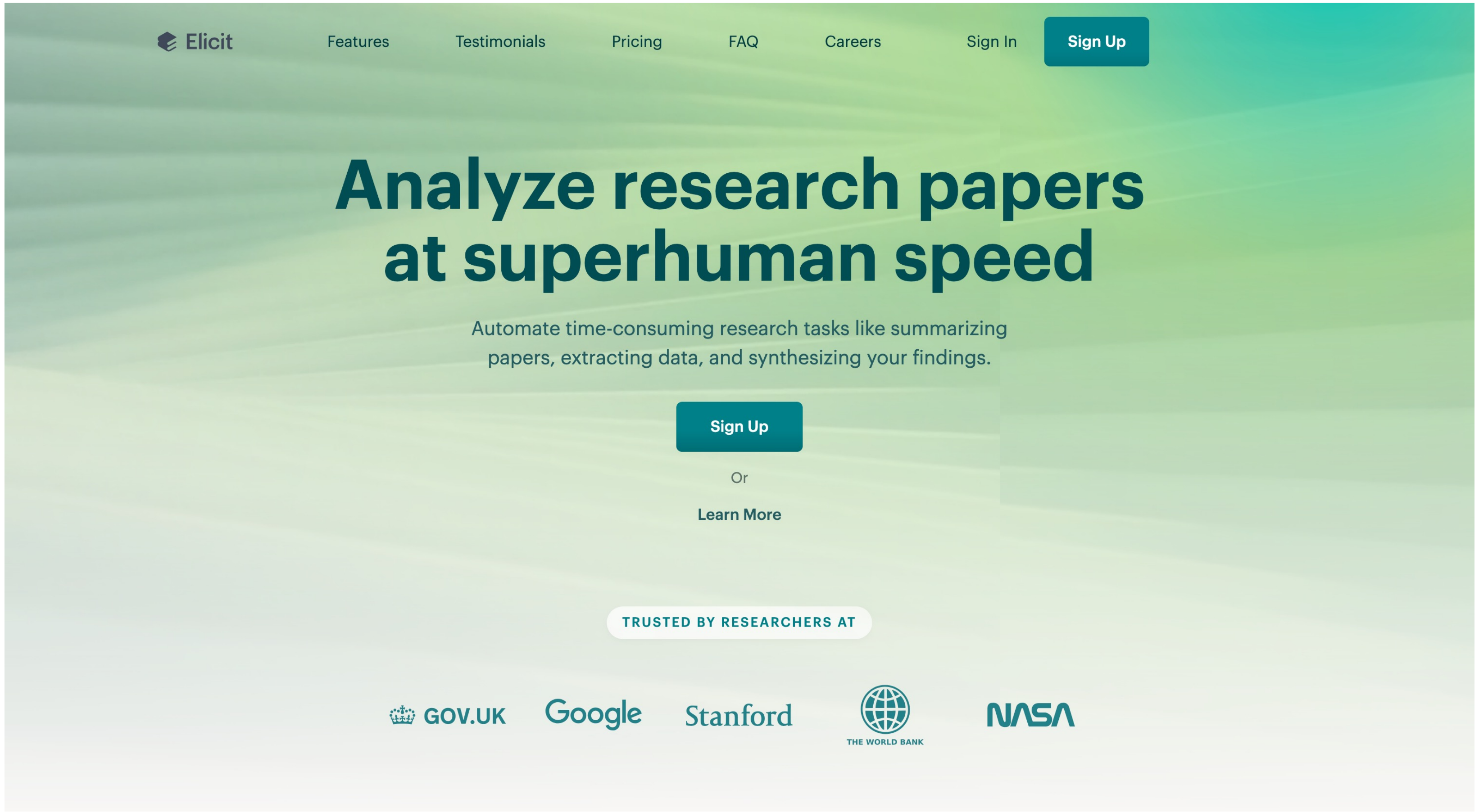
- **Perplexity**[1] might be a good starting point to search your hypotheses or ideas



Sources: [1] https://www.perplexity.ai/

Literature and State of the Art Review

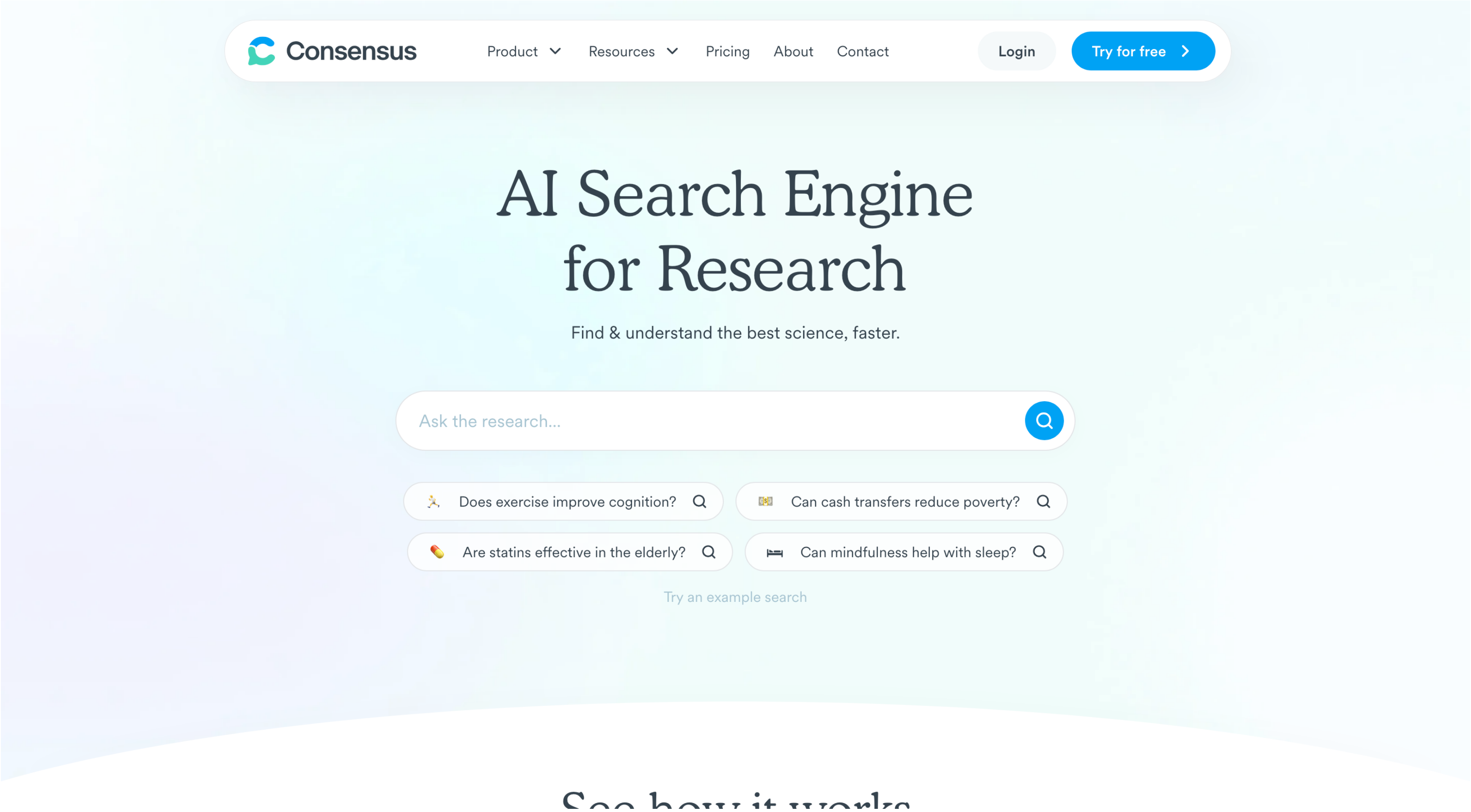
- **Elicit**[1] allows researchers to get a list of relevant references/articles/papers related to the question they ask in the platform



Sources: [1] https://elicit.com/welcome

Literature and State of the Art Review

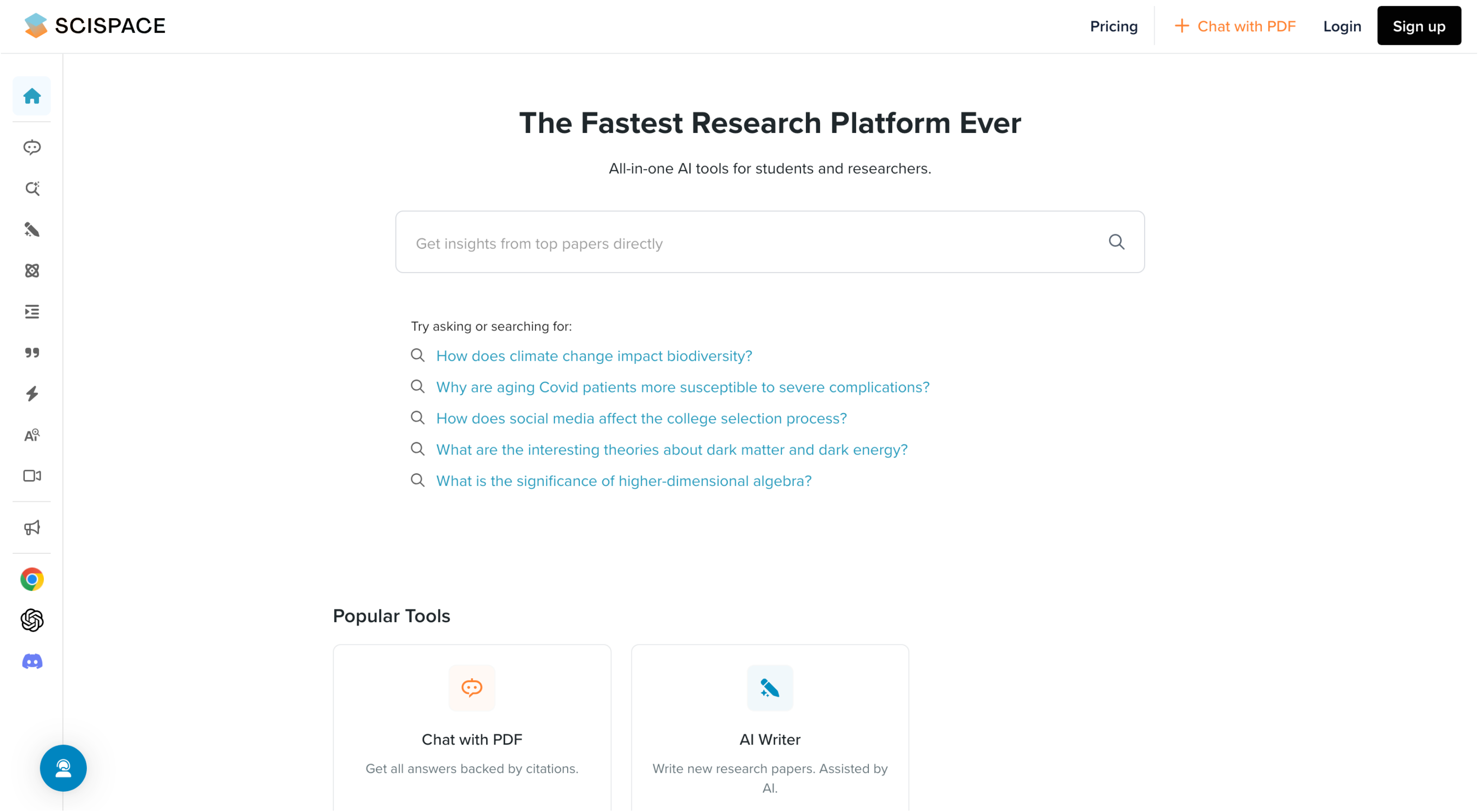
- **Consensus**[1] helps academics to get references and appropriate literature related to their questions



Sources: [1] https://consensus.app/?home=true

Literature and State of the Art Review

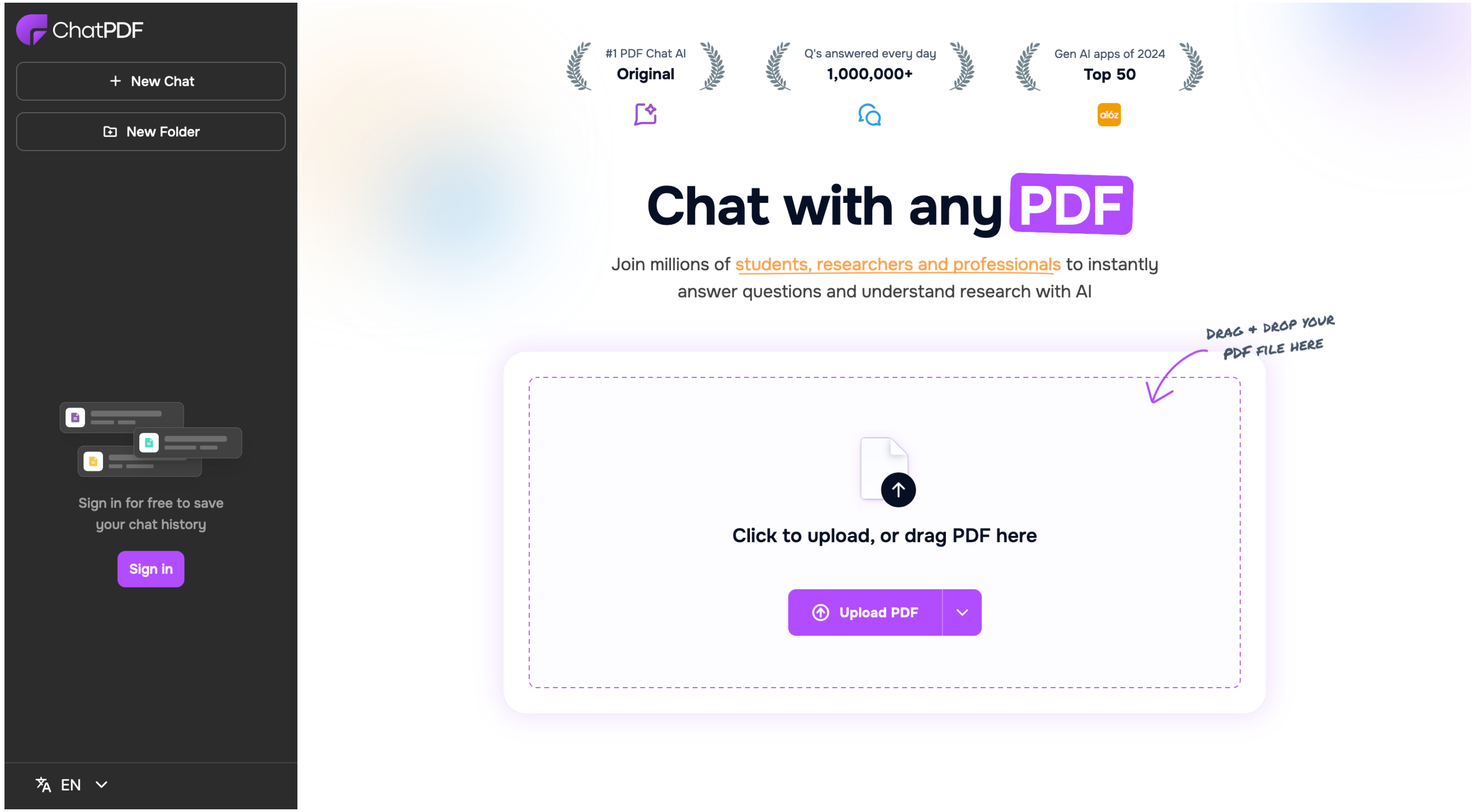
- **Scispace**[1] aims to help researchers to understand research papers better



Sources: [1] https://typeset.io/

Literature and State of the Art Review

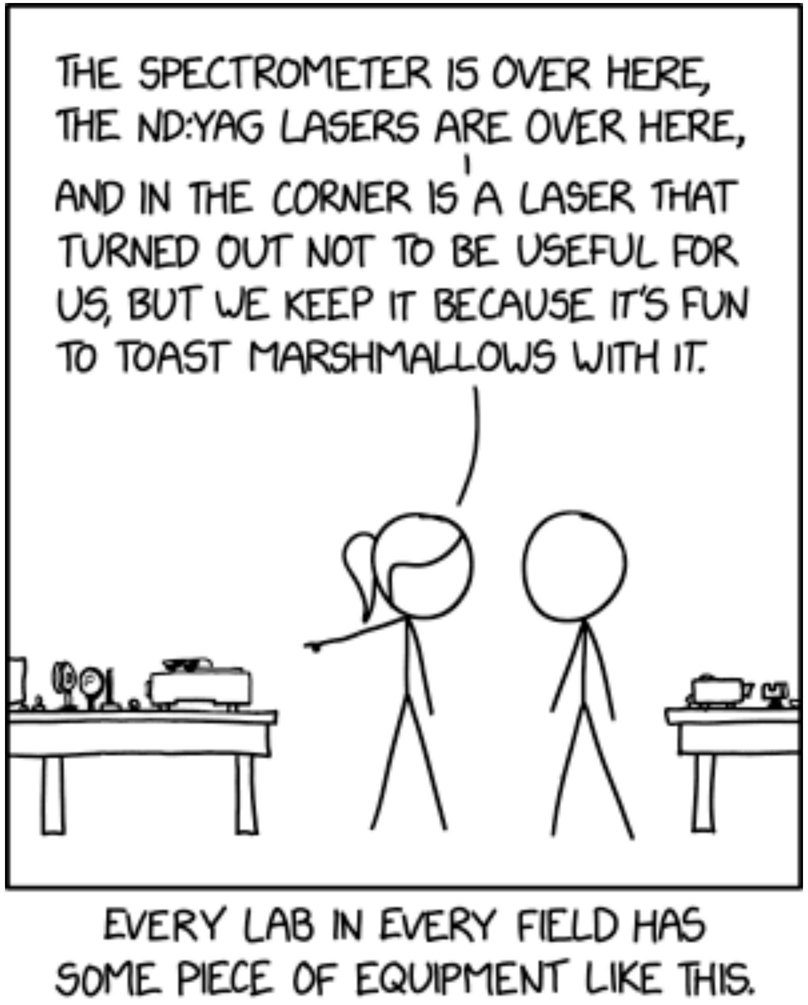
- **ChatPDF**[1] lets you chat with any PDF



Sources: [1] https://www.chatpdf.com/

Experimental Protocol, Materials and Methods

- Sometimes, you can’t use AI…

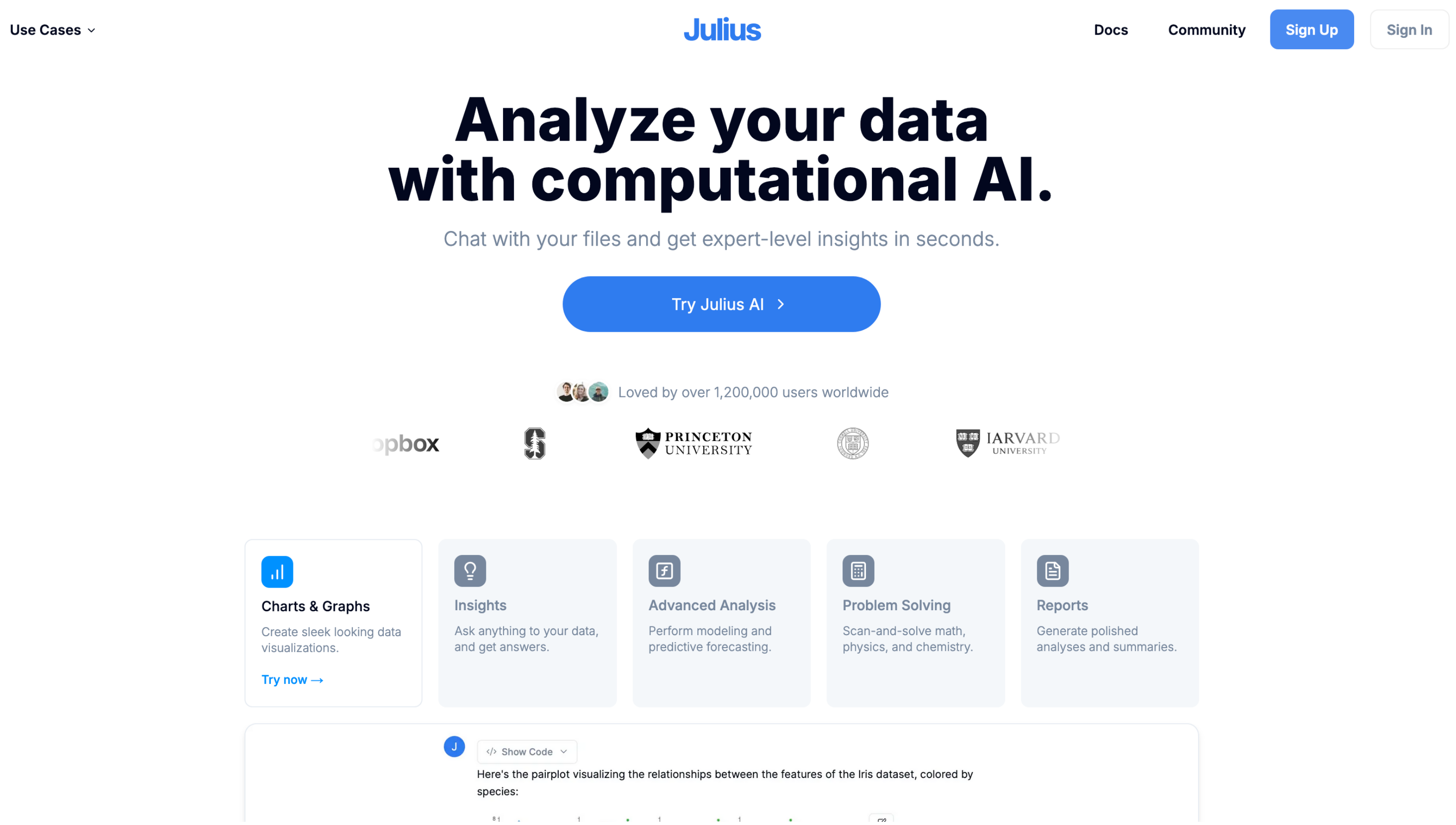


Sources: xkcd "2514: Lab Equipment"

INESCTEC

Conduct Experiments, Generate and Analyse Data

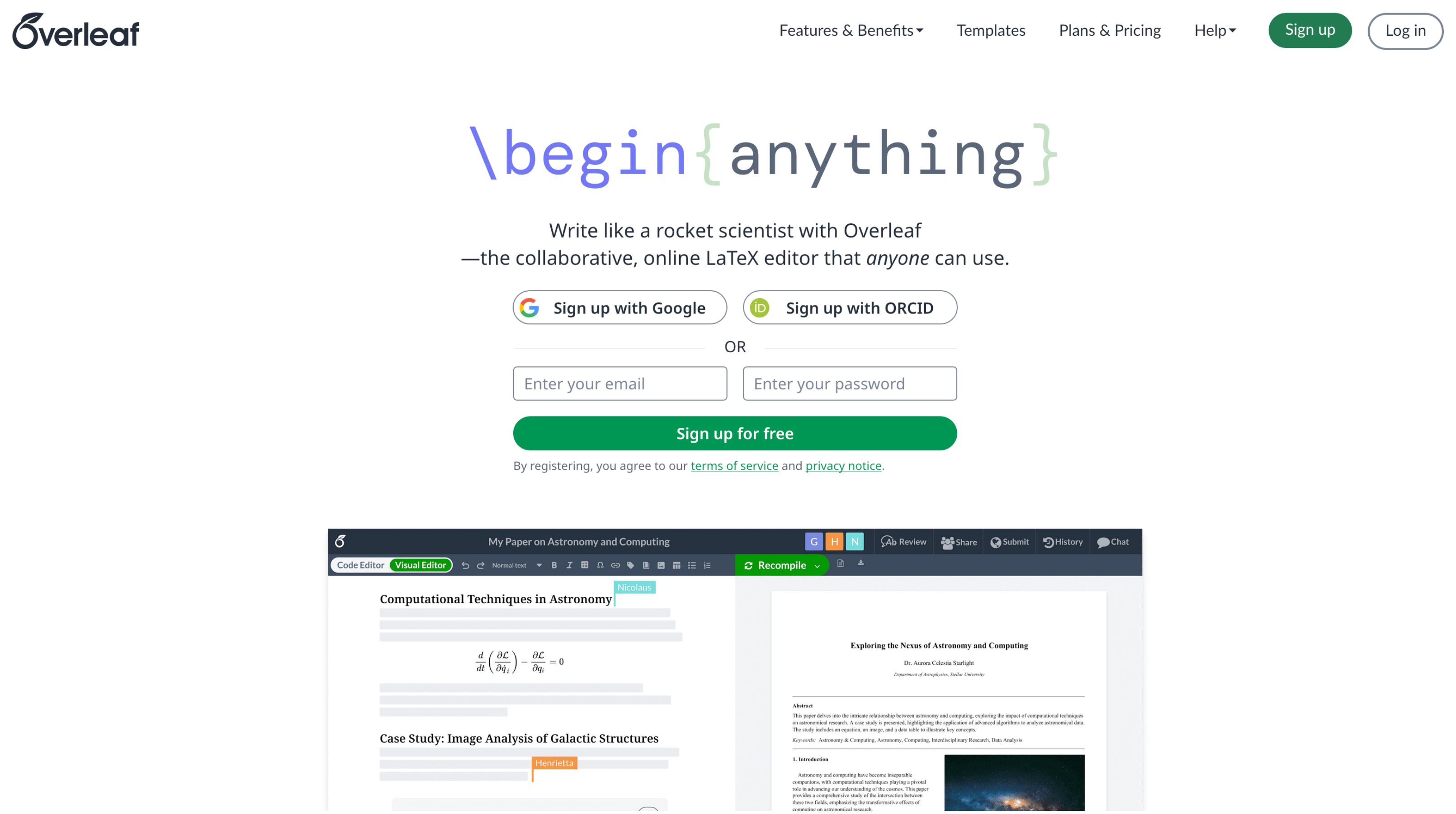
- **Julius**[1] intends to help you with data analysis



Sources: [1] https://julius.ai/

Write a Report, Dissertation, Thesis or Paper

- **Overleaf**[1] is not about AI (although it has some AI running in the background), but it will (probably) be your best friend during the writing process

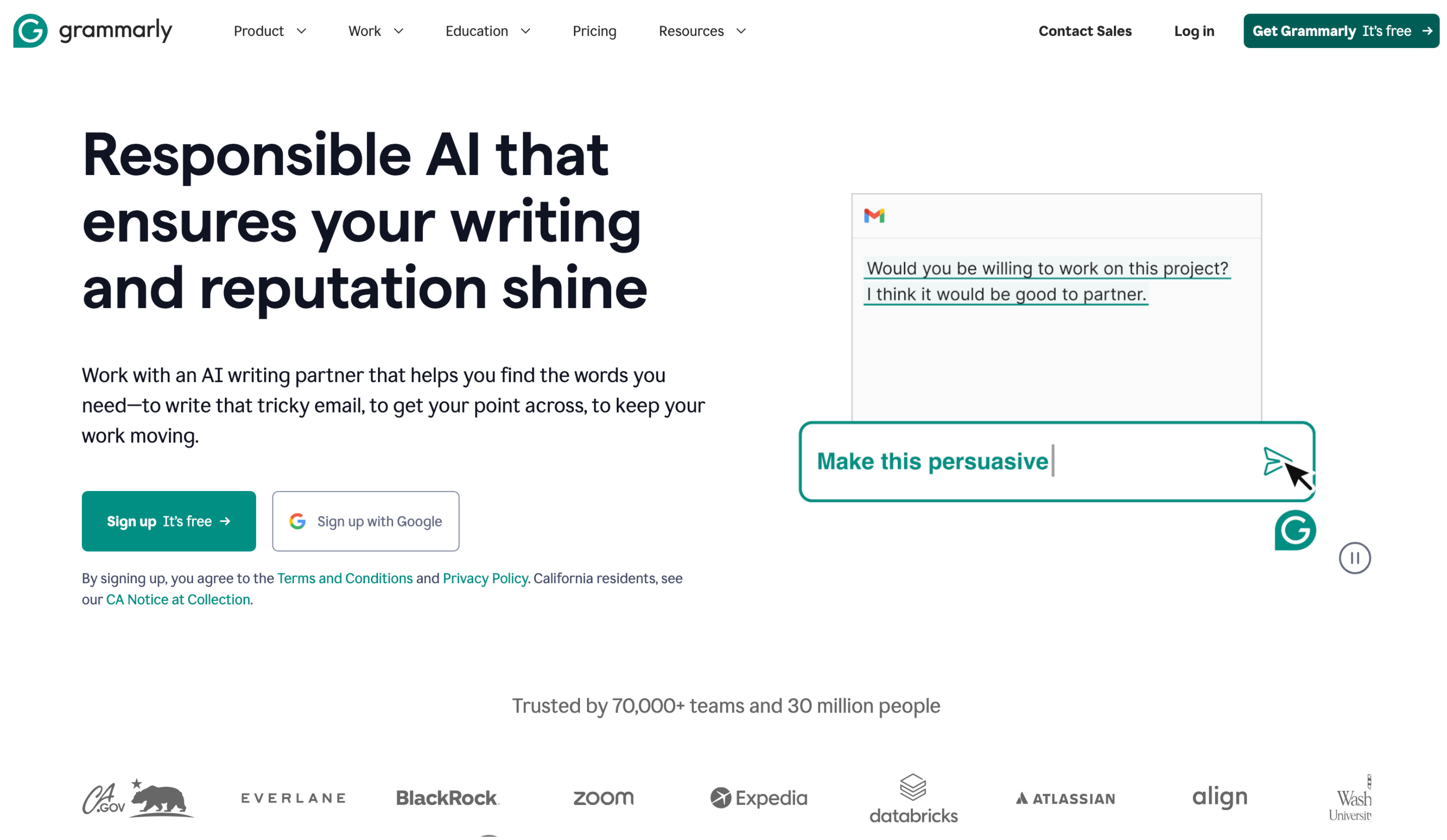


Sources: [1] https://www.overleaf.com/

INESCTEC

Write a Report, Dissertation, Thesis or Paper

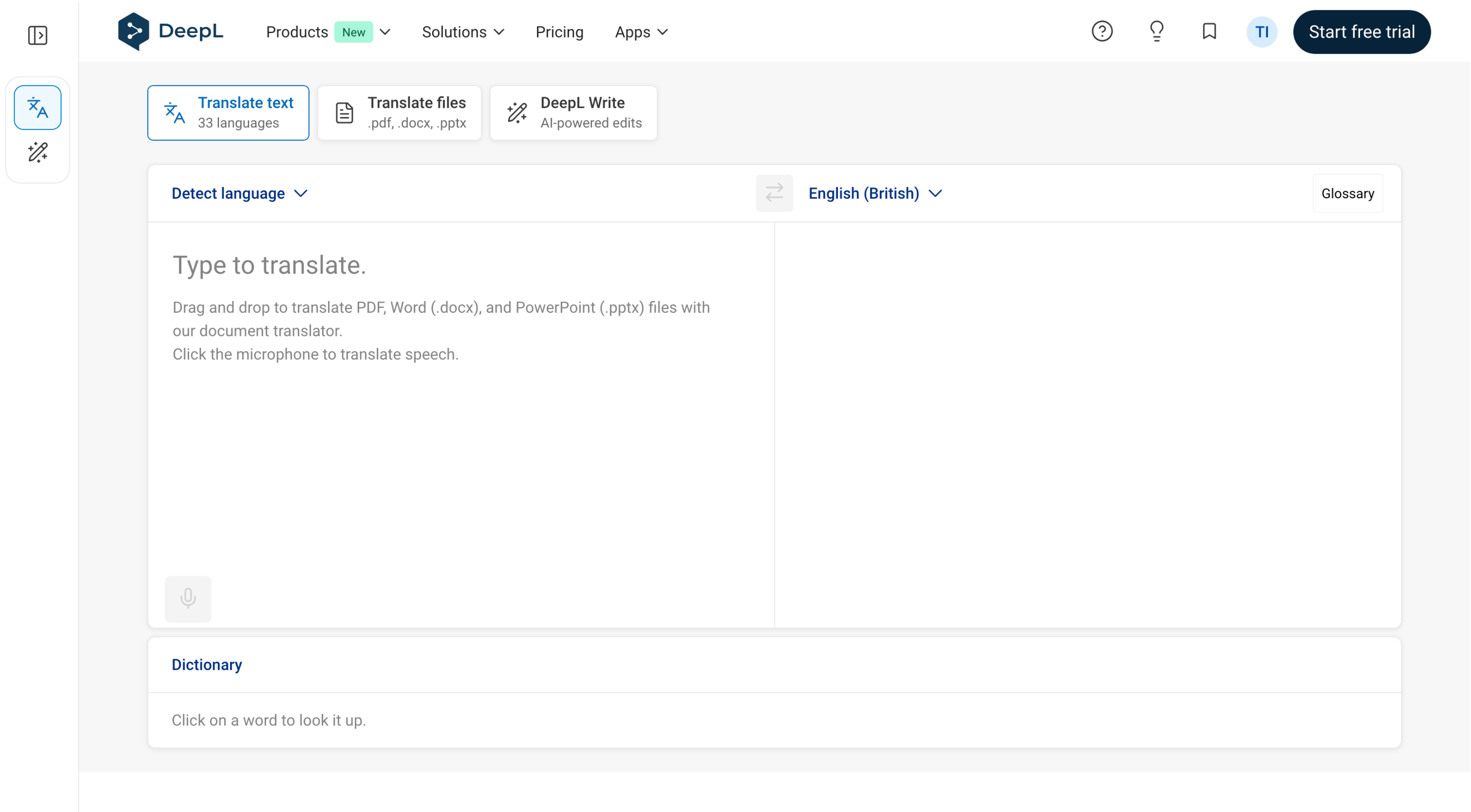
- **Grammarly**[1] uses AI to help you write better



Sources: [1] https://www.grammarly.com/

Write a Report, Dissertation, Thesis or Paper

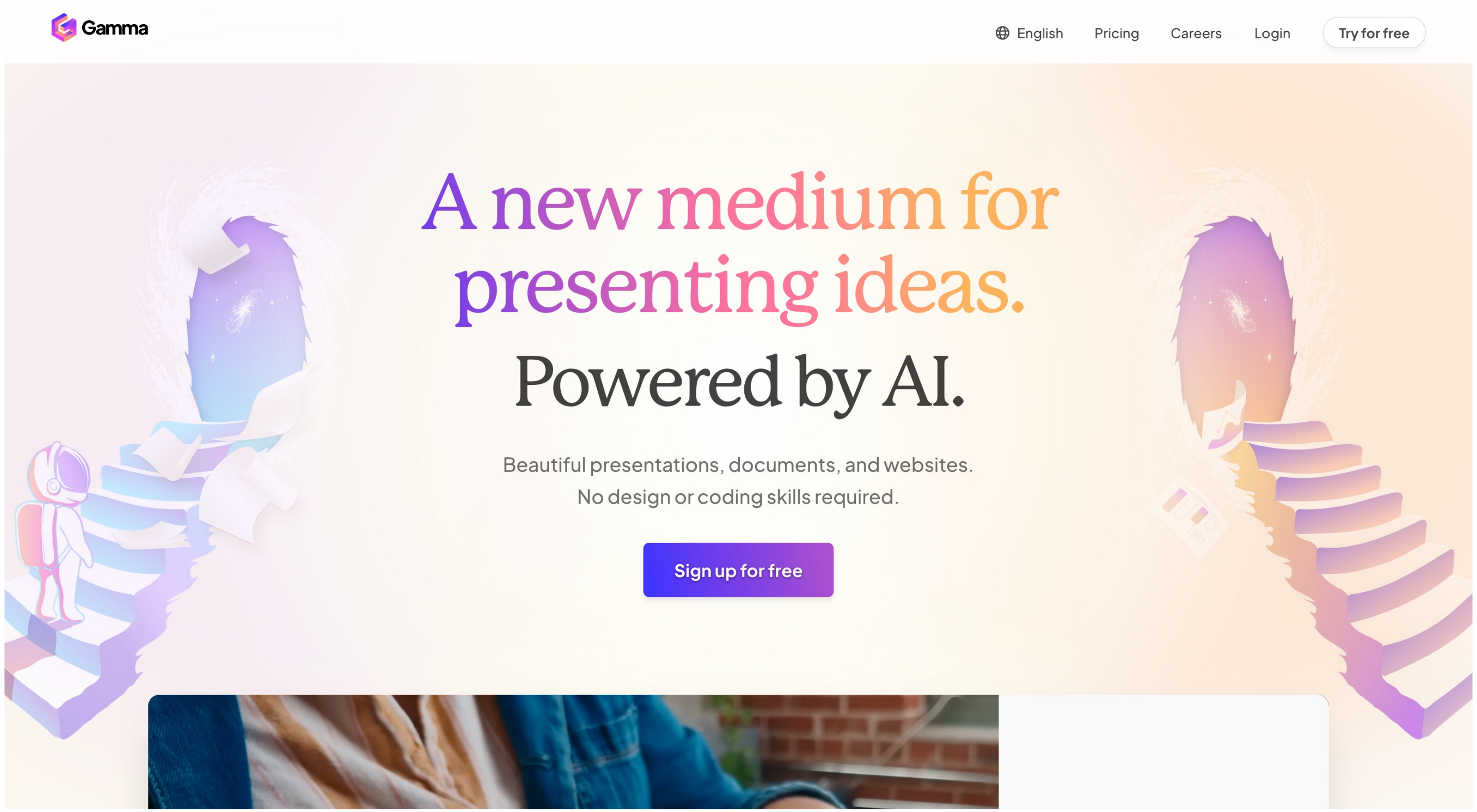
- **DeepL[1]** will help you to translate the difficult stuff



Sources: [1] https://hnd.www.deepl.com/en/translator

Publish and Present the Scientific Findings

- **Gamma[1]** might give you some inspiration when creating your presentations



Sources: [1] https://gamma.app/

4. Surfing the Waves of AI: Take-home Messages

Can we trust AI? Towards “Trustworthy AI” in the EU

INESCTEC

- The European Commission appointed a group of experts to provide advice on its artificial intelligence strategy: **High-Level Expert Group on AI**[1]

- **According to the Guidelines, trustworthy AI should be:**
  - **Lawful**: respecting all applicable laws and regulations
  - **Ethical**: respecting ethical principles and values
  - **Robust**: both from a technical perspective while taking into account its social environment

- Several important **guidelines were proposed**[2]:
  - **Human agency and oversight**: AI systems should empower human beings
  - **Technical Robustness and safety**: AI systems need to be resilient and secure
  - **Privacy and data governance**: data governance mechanisms must be ensured
  - **Transparency**: the data, system and AI business models should be transparent
  - **Diversity, non-discrimination and fairness**: AI systems should be accessible to all
  - **Societal and environmental well-being**: AI systems should benefit all human beings
  - **Accountability**: ensure responsibility and accountability for AI systems and their outcomes

Sources: [1] https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai, **[2]** https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

The AI Act and how it will impact our lives

- The AI Act[1] is a **document proposed by the European Commission** that contains several **harmonised rules**[2] regarding **AI applications**, emphasising that its approach is shaped by EU values and **risk-based**, ensuring both **safety** and **fundamental rights protection**

- What does the AI Act propose?[2]
    - **Prohibition of unacceptable AI practices** (e.g., social scoring)
    - **Regulation of high-risk AI systems** (e.g., AI used in the context of recruitment)
    - **Conformity assessment** (i.e., under the EU product safety framework)
    - **Transparency obligations for potentially deceptive AI systems**
    - **Ex post market surveillance** (i.e., post-market monitoring system)
    - **Governance** (i.e., authorities must be appointed for the application and implementation)
    - **Pre-emption of national AI regulatory frameworks** (i.e., regulated by the EU)
    - **Monitoring and enforcement** (i.e., done by the Member States)
    - **Compliance with the prohibitions and regulatory requirements**

Sources: [1] https://www.eipa.eu/publications/briefing/the-artificial-intelligence-act-proposal-and-its-implications-for-member-states/, **[2]** https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

INESCTEC

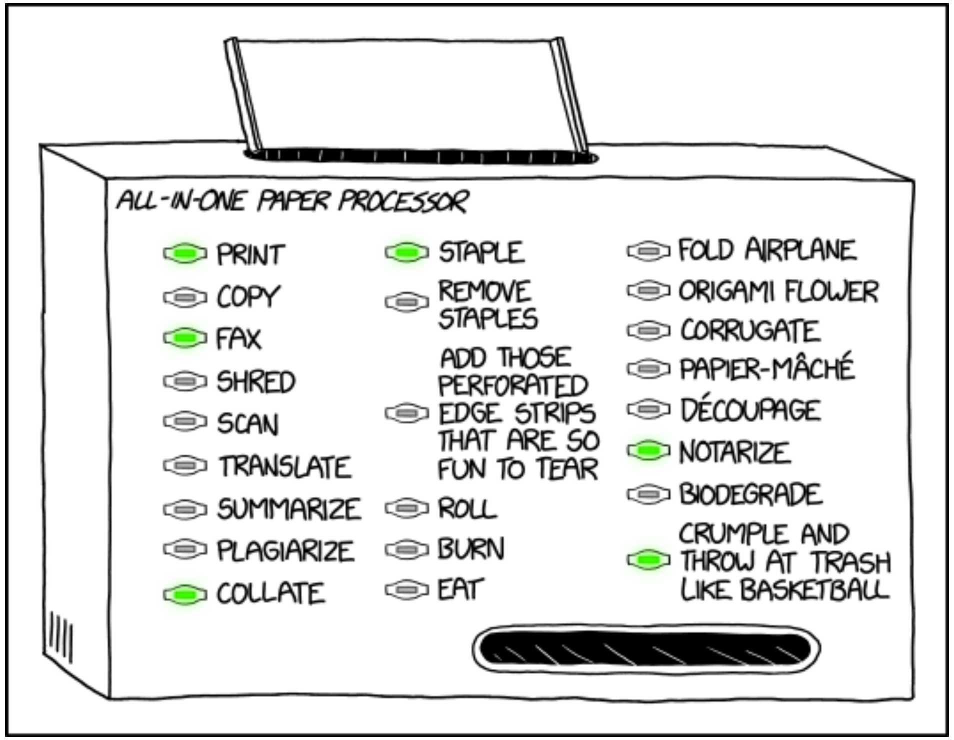
An Accurate and Honest Summary of this Session

- **The development of data-driven artificial intelligence applications is impacting our lives**, motivating the need for ethical, legal and technical regulatory frameworks based on specific principles: **accountability**, **interpretability**, **fairness**, **safety**, **privacy**

- **Think about AI software and applications as tools:** worry about knowing how these algorithms work and how you can leverage their power to improve the quality of your research and work

- **Always bear in mind the principles of Ethics and Research Integrity**, and think of the impacts of using AI-driven applications in your research and work

- **Multidisciplinary work** is, more than ever, of utmost importance and useful



Sources: xkcd "2369: All-in-One"

Artificial Intelligence Applied to Scientific Research and Ethics: How can we leverage AI algorithms to achieve better science?

Mestrado em Energias Sustentáveis | Workshop

December 10, 2024 | Instituto Superior de Engenharia do Porto

Tiago Filipe Sousa Gonçalves

https://tiagofilipesousagoncalves.github.io/

INESCTEC

INSTITUTE FOR SYSTEMS AND COMPUTER ENGINEERING, TECHNOLOGY AND SCIENCE