Hpi Workshop Helsingor, 2024

AI in Science: SWOT analysis

Strength, Weaknesses, Opportunities, Threats

Important Definitions:

- “Generative“ AI: Large language models (LLM), natural language processing (NLP) such as ChatGPT,

Translations, music/voice generators, image/movie generators

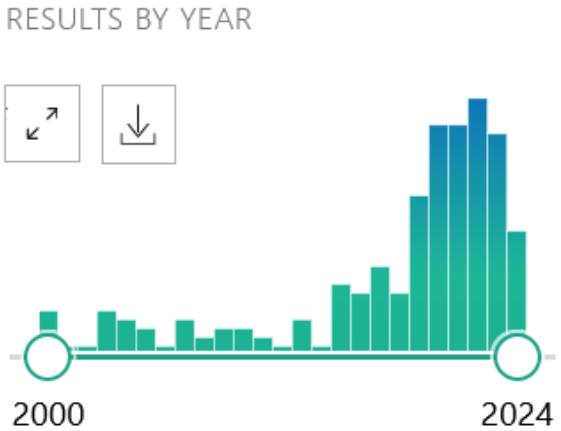
(statistics) Closed/Blackbox (original data not known, usually based on large training datasets; strong “admixture“ = mosaics texts)

- Logical and logical data-based AI: alignments, -Omics data analysis, evolutionary distance analysis, structural models (alphafold)

More “open box“ (original data known and accessible), includes machine-learning algorithms

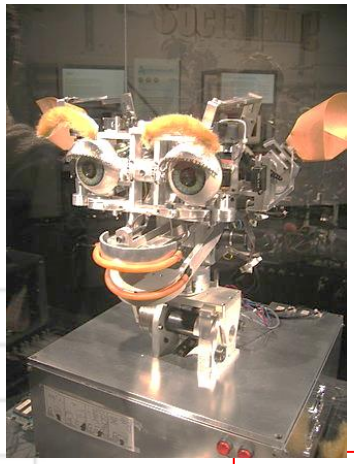
H. pylori: 164 papers, mostly logical AI use

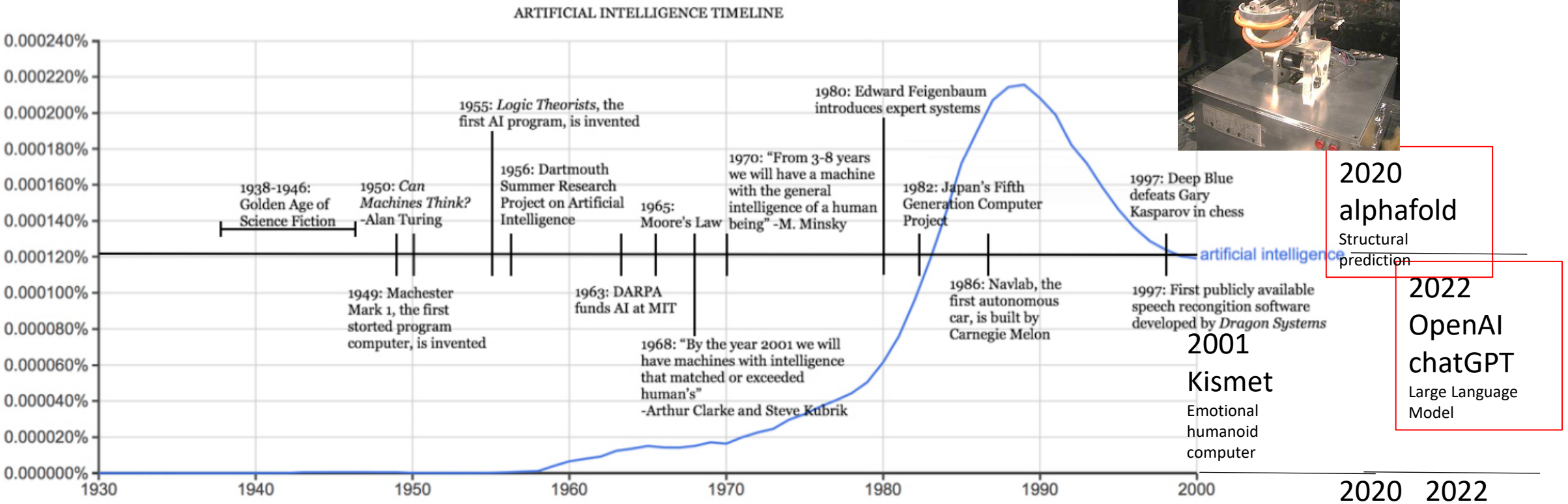
(clinical, meta reviews)



History:

- Old models and algorithms (<30 years); current major difference is the amount of training data going in and the increased calculation speed and storage capacity of large supercomputers and computing networks





Generative AI (ChatGPT), example

- PROMPT: The study of **Helicobacter pylori** (H. pylori) has advanced significantly, but several important questions remain open:

1. **Mechanisms of Pathogenesis**:
- - How exactly does H. pylori cause gastric cancer and other diseases? Knowing detailed molecular pathways and interactions between the bacteria and host cells is crucial.
2. **Antibiotic Resistance**:
- - How can we effectively manage and overcome the increasing AMR in H. pylori strains? Alternative therapies?
3. **Immune Response and Host Interaction**:
- - How does H. pylori evade the host immune system? Role of host genetic factors?
4. **Transmission and Epidemiology**:
- - What are the precise routes and mechanisms of H. pylori transmission? Public health interventions?
5. **Microbiome Interactions**:
- - How does H. pylori interact with the rest of the gastric and intestinal microbiota?
6. **Vaccine Development**:
- - What are the challenges for effective vaccine against H. pylori and how can this be solved?
7. **Strain Variability and Virulence Factors**:
- - How do different H. pylori strains vary in terms of virulence and disease-causing potential?
8. **Environmental and Lifestyle Factors**:
- - What environmental and lifestyle factors influence the risk of H. pylori infection and the severity of associated diseases?
9. **Long-term Consequences of Infection**:
- - What are the long-term health consequences of H. pylori infection beyond gastric diseases?
10. **Diagnosis and Screening**: - How can diagnostic methods and screening be improved to detect H. pylori (non-invasively)

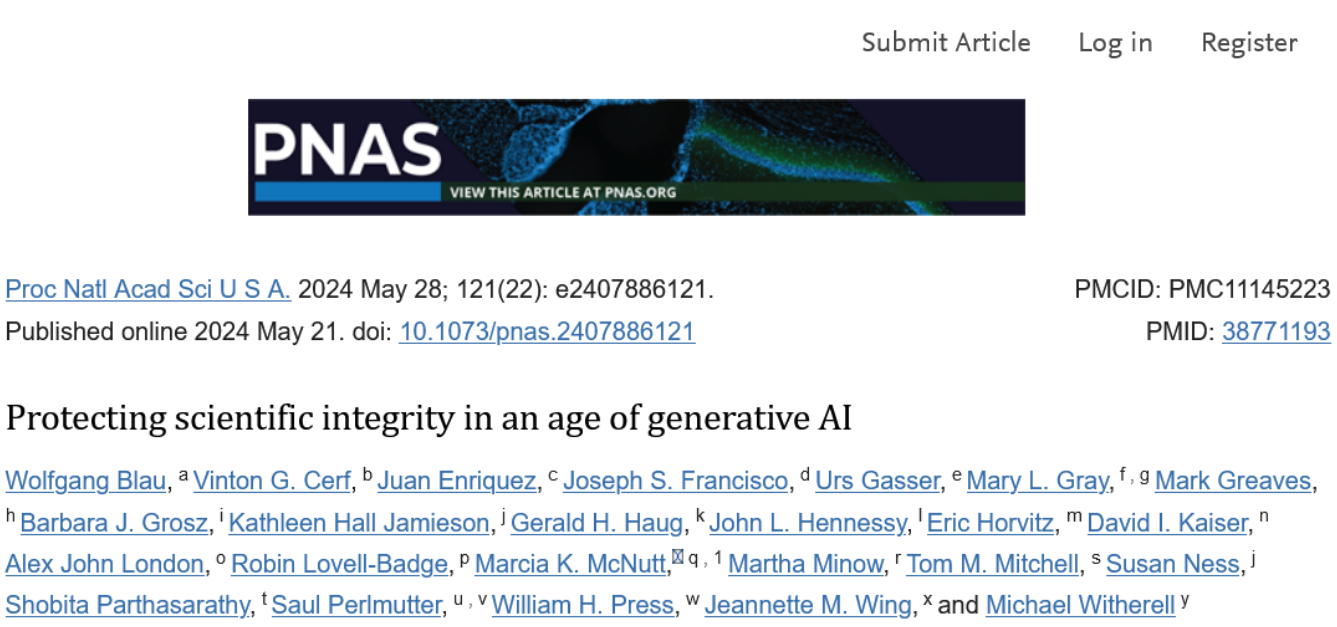
Addressing these questions requires interdisciplinary research combining microbiology, immunology, genetics, epidemiology, and

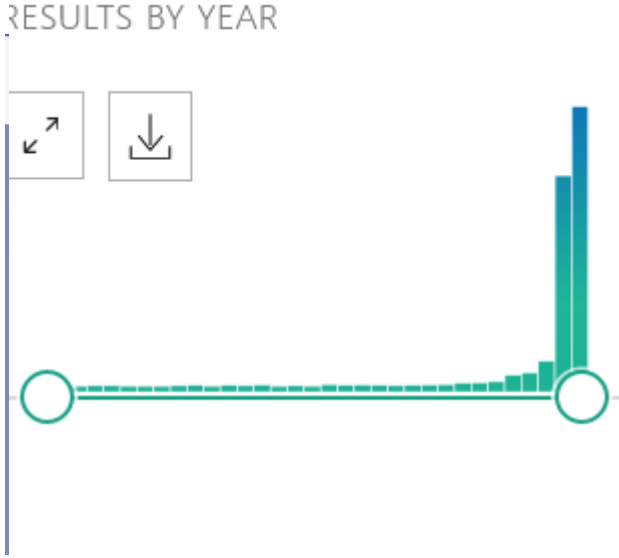
SWOT – generative AI (G), logical AI (if x- then y; (L))

Strengths (Science):	**Opportunities (Science):**
Can possibly be time-saving (G, L) **Analysis of large datasets, publication collections (G)** **Facilitate all data analysis (e.g. Omics) (L)** **Facilitate structural discovery (proteins, others molecules (L)** **Facilitate drug discovery (L)** **Failitate image analysis and medical image analysis (e.g. pathology)**	**Improve output** **Improve detectability of mistakes/plagiarism** **Provide open code**
Weaknessess (Science, General): (growing) Business models behind it (G, L) Often no sources given (G) "It is a mix" Sources not transparent (G) No open code available (G, L) May support or induce plagiarism (G) May infringe copyrights/confidentiality (G) May contain filters/biases (G) Degeneration and mistakes of output and code (G) Appropriation of content/data/art created by others (G, L) [May inhibit intellectual training]	**Threats** (to Science, but also to the models themselves): Degeneration of training data (G) (vicious cycle) Degeneration and mistakes of output and code (G) Faulty data, faulty interpretation (garbage in, garbage out) Large energy and resource expenditure (G, L) 100 households-1 d, 4 Mio liters water for pre-training ChatGPT

Current Policies/Measures

- Recent papers for use guidelines of generative AI in science/medicine (NIH authors)





- Journal policies: allow no use of AI (LL) for text generation or without precisely giving sources; only for text polishing; humans remain responsible for content; no use of AI (LL) for generating manuscript reviews; give AI use for selected purposes. No use for image generation

- Funding agency policies (similar to journals); no use of AI (LL) for making reviews

- More SWOT
- More comments
- Legitimate/ GSP Uses