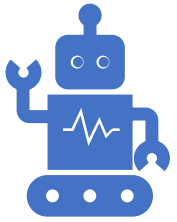
Evolution of Science in the AI and Big Data Era: Modeling, Replication, Expertise

Bartłomiej Nawara



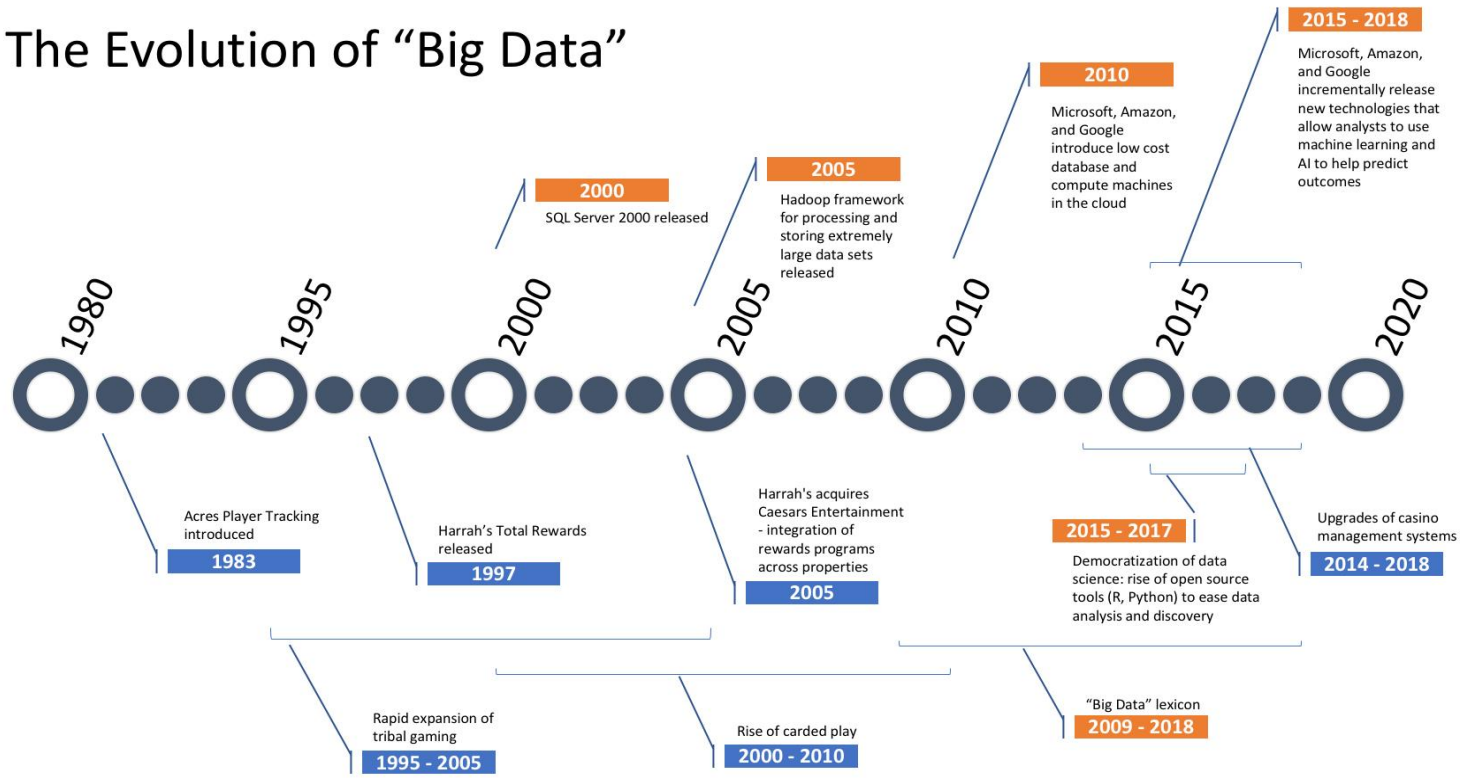
Introduction

Two parts of the presentation

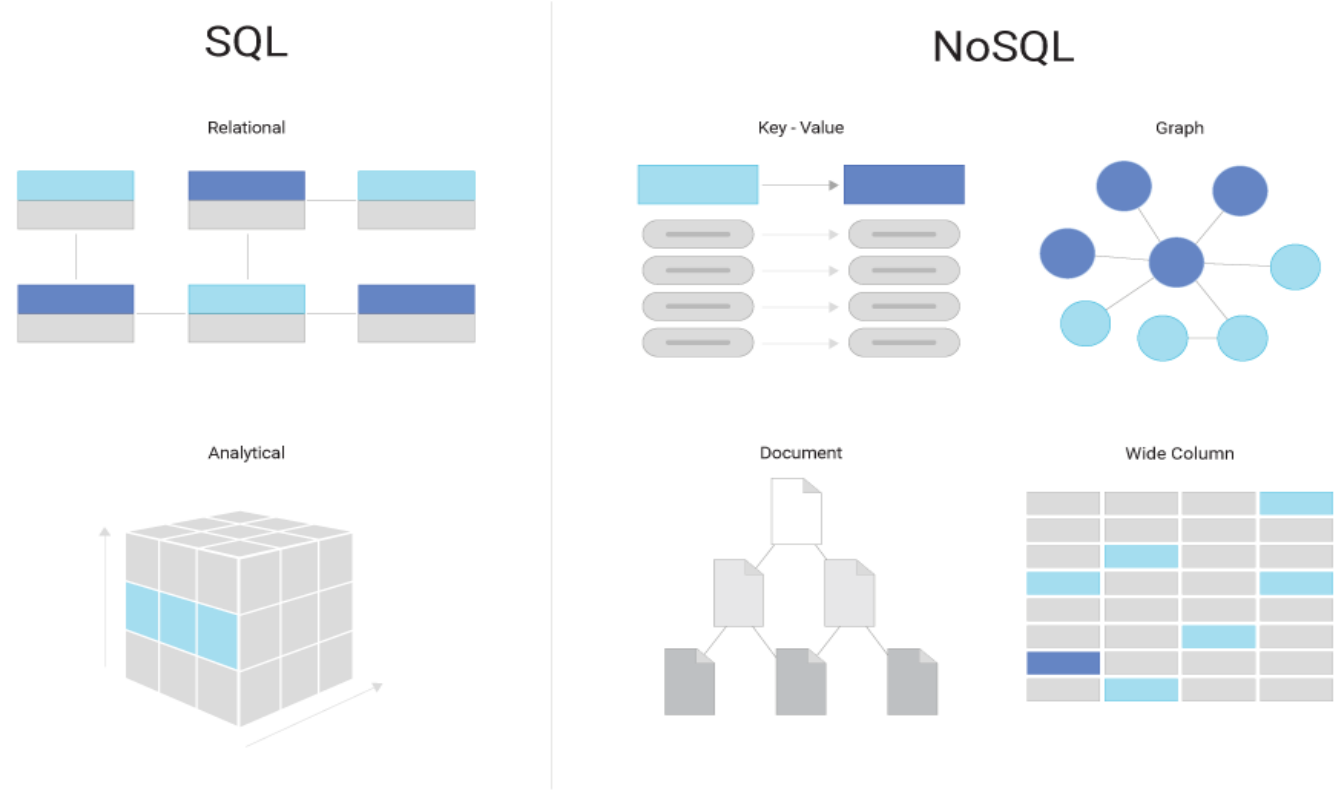
Technological aspects of Big Data/AI Science

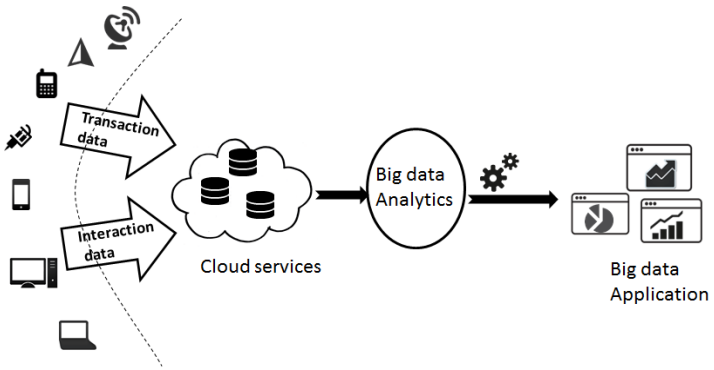
Philosophical and methodological aspects of Big Data/AI Science

The Rise of Big Data Science



From SQL to NoSQL Databases





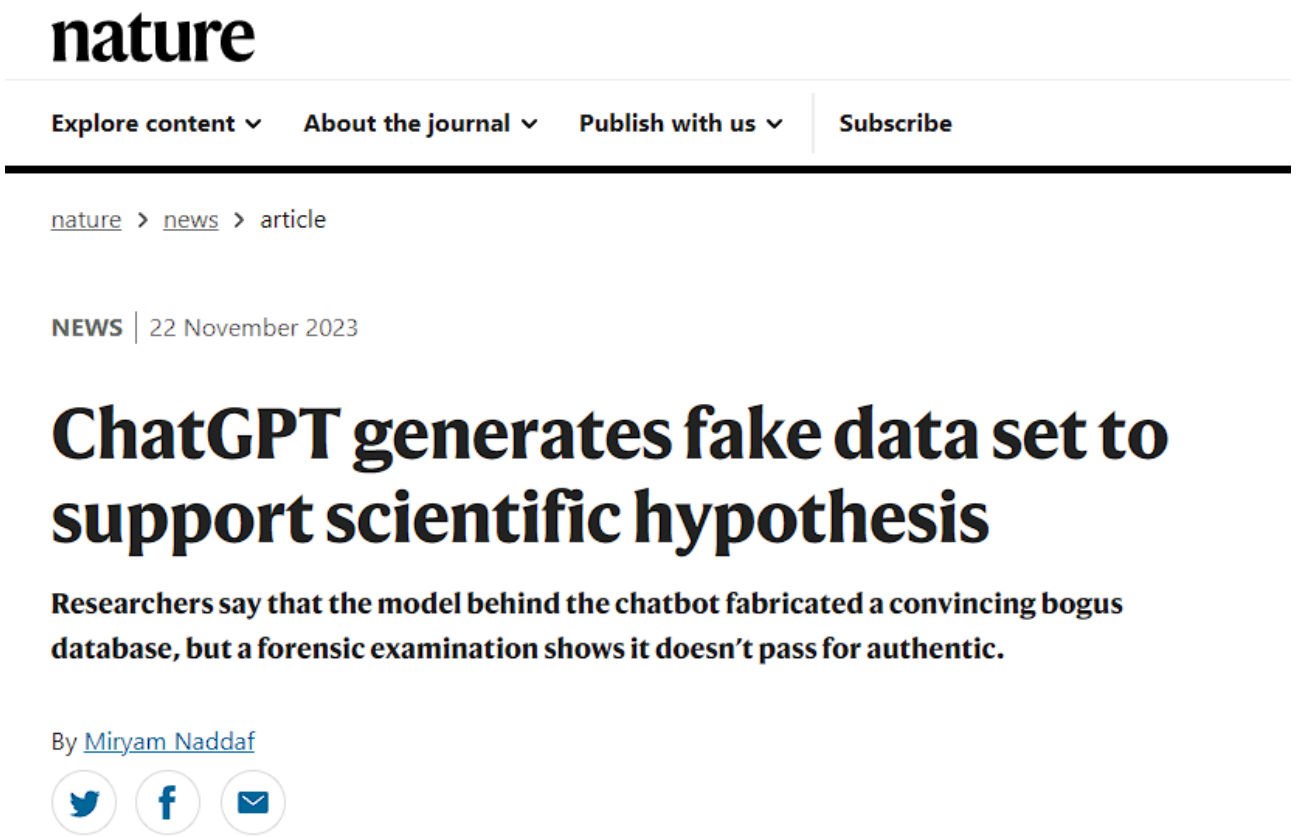
Cloud Solutions as one of central points of the revolution

- Flexible and Scalable Infrastructure
- Cost-Effectiveness and Accessibility
- Integrated Advanced Analytics Tools
- Security, Compliance, and Speed

LLMs Transforming BDS Research Methods

- Enhanced Data Analysis and Interpretation – f.e. Advanced Data Analysis via Chat GPT
- Data Cleaning and Preprocessing - LLMs aid in automating the process of cleaning and organizing data, which is a fundamental step in BDS (WARNING – Potential of fake dataset and fake results – example in the next slide)
- Automated Content Generation - generate research summaries, literature reviews, or even draft research papers based on existing data – f.e. Scite.AI
- Many more to come

Generating fake dataset to suport faked scientific hypotheses



nature

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

nature › news › article

NEWS | 22 November 2023

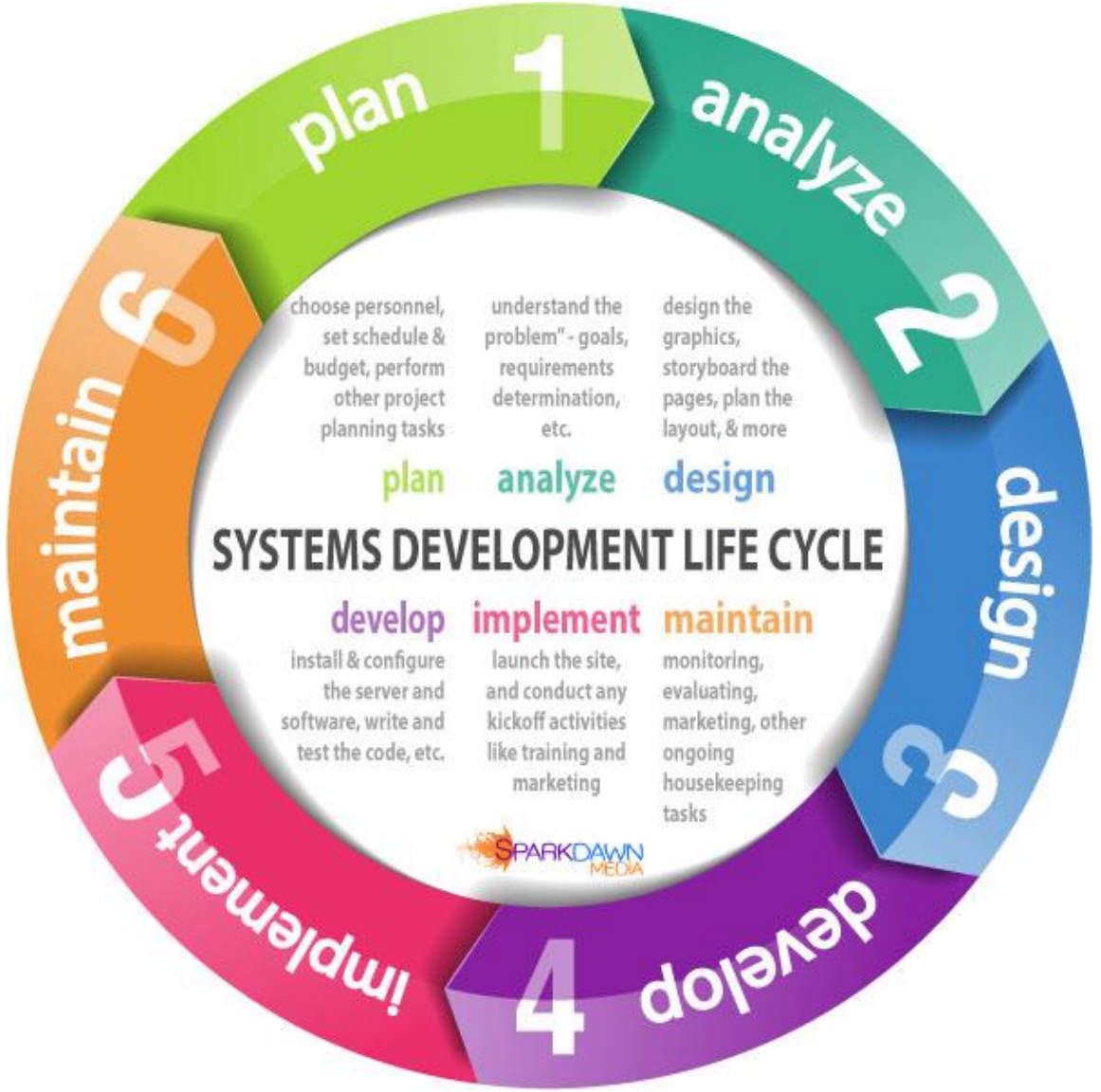
ChatGPT generates fake data set to support scientific hypothesis

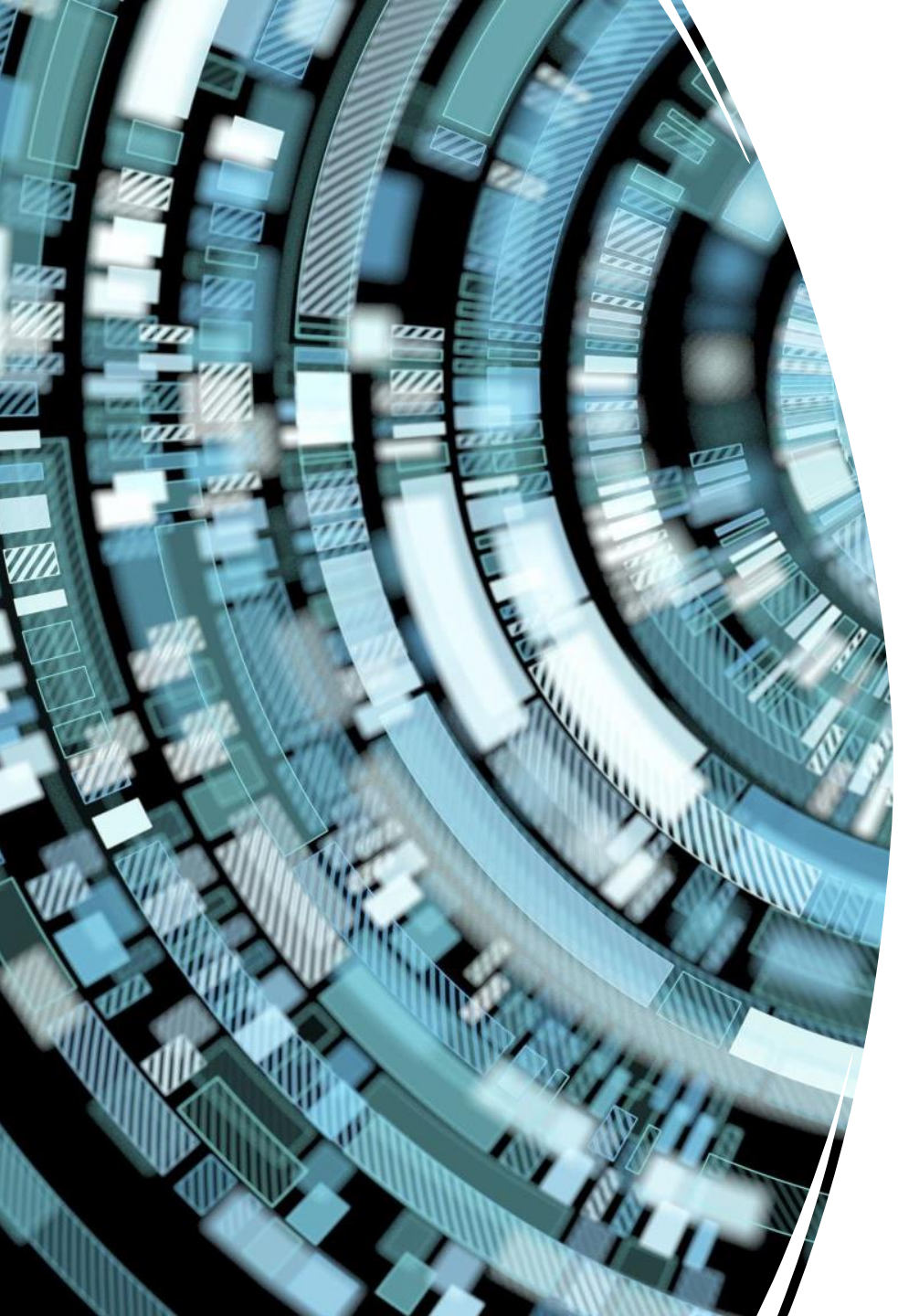
Researchers say that the model behind the chatbot fabricated a convincing bogus database, but a forensic examination shows it doesn't pass for authentic.

By Miryam Naddaf

Philosophical Dimensions of Big Data Science

Drawing parallels with the Systems Development Life Cycle (SDLC), BDS is seen as a methodology deeply ingrained with software processes, highlighting the significance of software in the evolution and execution of scientific research in the age of BD.





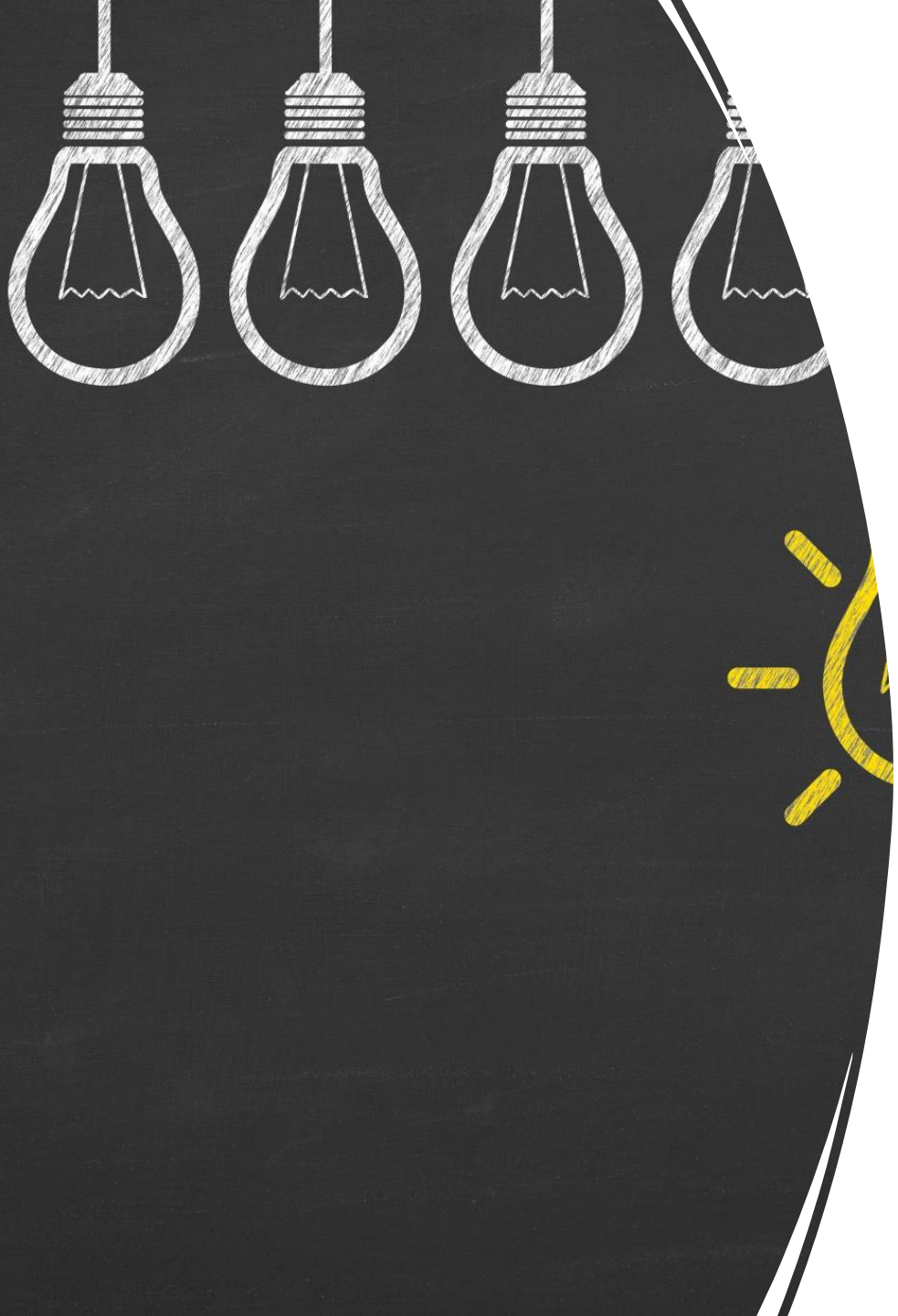
Software-Intensive and Software-Ladent Science

- Automated Inferences from Software-Intensive Science
- Automated reasoning or automated hypothesis creation?
- The role of software and hardware
- Software Complexity - numer of software path based on conitional instructions vs no direct programming in AI models
- Nautre of modeling in Big Data (and AI) Science – deep hierarchical modeling in traditional sciences vs narrow domains, poor generalisation abilities

Software-Ladent Science

- Big Data Science as **'software-laden' (SL)** science—a term that implies an inextricable dependency on software to generate results. Big Data Science not only utilizes software but is fundamentally based on it.

- SL Science **lacks of epistemic transparency** - methods employed in Big Data Science are often epistemically opaque in the sense defined by Humphreys, where an agent may not know all epistemically relevant components of a process. This opacity, often due to the explainability challenges of complex AI systems, does not make these methods intrinsically epistemically opaque. Instead, the opacity stems from the difficulty in explaining the internal workings due to numerous parameters and the complexity of algorithms.

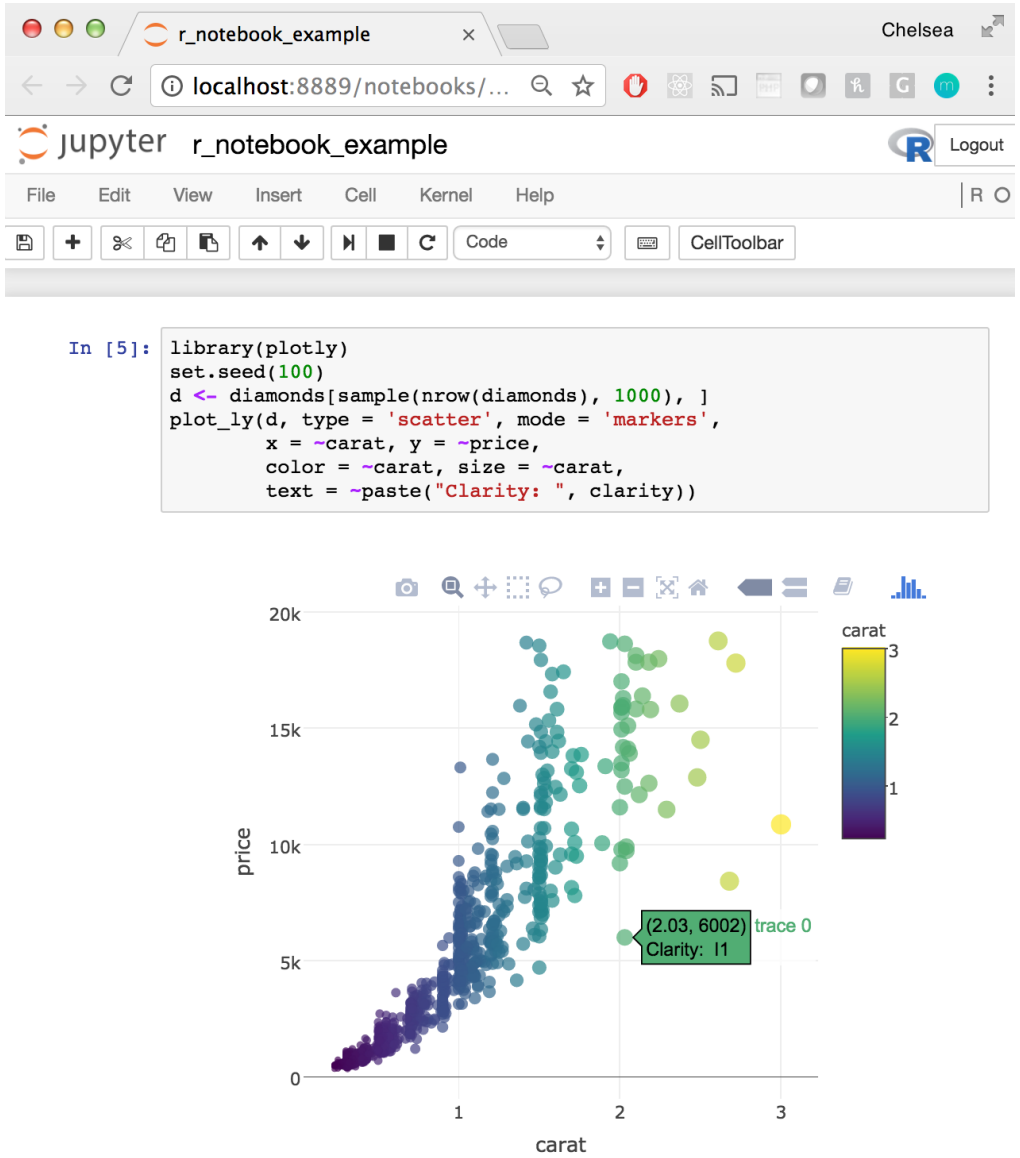


Scientifical implications: replication and explainability

- Big Data Science has proven to be cognitively invaluable, delivering scientific results where traditional scientific methods fall short. **It is too significant in its contributions to be excluded from what is considered scientifically valid**.
- However, **Big Data Science does not fully meet some of the traditional standards expected of scientific practice**, particularly in terms of the replicability and explainability of its methods.

Replicability in AI and BDS Research

- Replication in Big Data Science is in fact the replication of the computational experiment.
- Can it be solved by the **Notebook method**? In some parts yes, in other no.



XAI and Redefining Expertise

- Unlike traditional AI methods, explainable AI (XAI) doesn't conform to deductive-nomological (DN) or causal-mechanistic (MP) models for explanations. XAI primarily operates within the framework of statistical relevance (SR) model, offering a different approach to understanding AI behavior.
- **Functional explanations** - focus on the practical aspects of explanation, considering factors like societal and scientific contexts. They redefine expertise in AI by emphasizing the ability to contribute to epistemic progress within a specific domain.

Summary - Challenges facing contemporary science in the era of Big Data and AI

- Scalability vs. Explainability - Larger models enhance scientific progress but are less comprehensible and harder to replicate.
- Incorporating New Methods - Science needs to adapt to effectively use and understand burgeoning, complex methodologies.
- Acceptance vs. Rejection - Deciding whether to embrace new, less transparent methods due to their benefits is a significant dilemma for the scientific community.
- Privatization of Research - The growing influence of private companies in AI research threatens to undermine the fundamental values of open and transparent science.

Thank you for your attention