

ROBHOOT

Automated Discovery-Knowledge Graphs in Federated Cooperation Networks v.1.0

April 28, 2020

Summary

Global sustainability is a major goal of humanity. Many studies have shown global sustainability could be achieved by strengthening transparency, feedbacks and rapid access to reproducible information among social, ecological, economical, technological and governance systems. Sustainability goals, however, strongly depend on global access to evidence-, and discovery-based knowledge gaps. Yet, science-enabled technologies targeting global knowledge gaps to reach sustainability goals are at a very incipient stage of development. We introduce data- and causal-knowledge graphs, the discovery-knowledge graphs, in federated networks for a sustainable- and knowledge-inspired society. Discovery-knowledge graphs running on a federated network encompass a hybrid-technology to lay out the foundation of an open- and cooperative-science ecosystem to automate discovery in global emergency and sustainability challenges. The project summarized here is not set out to deliver automated discovery-knowledge graphs in federated networks, but to provide the architecture of a science-enabled technology, as a proof-of-principle, to connect global human sustainability challenges to knowledge-inspired societies.

1 Excellence

1.1 Radical vision of a science-enabled technology

- Describe the vision of a radically-new science-enabled technology that the project would contribute towards
- The project will contribute towards automated discovery-knowledge graphs in cooperative federate networks targeting robust and rapid discovery in the face of rapidly emerging global sustainability challenges.
- Describe how this vision surpasses substantially any technological paradigms that currently exist or are under development.
- Technological paradigms for discovery targeting global sustainability are currently based on highly fragmented, non-reproducible, and competitive technologies. Discovery-knowledge graphs in cooperative networks instead focus on the need of rapidly responding to global sustainability challenges emphasizing cooperation and reproducibility.
- Describe the overall and specific objectives for the project, which should be clear, measurable, realistic and achievable within the duration of the project. (The details of the project plan belong to the Implementation section).
- *ROBHOOT* will be developed in four stages, each containing measurable and achievable goals, work packages, milestones and deliverables (Figure 2).

We are in the midst of the fourth industrial revolution, a transformation revolving around data driven intelligent machines. Yet, despite the rapid evolution of the digital ecosystem around data driven machines, discovery technologies facilitating global access to fully reproducible reports when solving complex governance, social, environmental and technological problems are particularly lacking (Figure 1 and Table 1) [1]. This lack of reproducible technologies is particularly relevant for targeting rapid information access in global emergency and sustainability challenges. How can automated and explainable (Table 1) data- and causal-knowledge graphs be integrated into discovery networks to generate predictive scenarios for global emerging challenges? How can discovery driven intelligent machines help to reach global sustainability goals? The *ROBHOOT* project integrates data- and causal-knowledge graphs, the discovery-knowledge graph, into intelligent networks to provide critical and accesible information to help humanity to take informed decisions in emergency and sustainability challenges (Figure 1 and Table 1).

More than half (i.e., 3.9 billion) of the global population is now online and using the Internet, which represents a more inclusive global information society. People are applying technology for good in powerful ways, from adopting decentralized technologies for humanitarian efforts, to improving agricultural practices and reducing waste in the global food supply chain [2]. Yet, current technological paradigm assisting humans for scientific inquiry is currently based on competitive schemes instead of intelligent global collaborative protocols in federated networks [3]. In addition, technologies for scientific inquiry are highly fragmented, partly solve reproducibility, are mostly developed in close-source software and contain many biases [4, 5, 6, 7, 8, 9, 10]. Thus, despite the importance of global access to discovery to close knowledge gaps for rapid information access in emergency and sustainability challenges, open-source technologies integrating discovery-knowledge graphs in federated, collaborative, intelligent networks are currently not in place. The goal of *ROBHOOT* is to propose a new hybrid-technology concept integrating data- and causal-knowledge graphs into intelligent networks to lay the foundation for a novel scientific discovery technology.

ROBHOOT will contribute towards discovery-knowledge graphs in intelligent networks to facilitate governance reproducible scenarios in rapidly changing global sustainability landscapes. *ROBHOOT* will be developed along four science-enabled technologies (Figures 1 and 2): *ROBHOOT v.1.0* will deploy question- and data-knowledge graphs for understanding bias and diversification of information sources when obtaining global data-architecture patterns (section 3.1). *ROBHOOT v.2.0* will integrate automated and explainable biology-inspired neural-networks to decipher causal-knowledge graphs from high-dimensional data (section 3.2). *ROBHOOT v.3.0* will explore decentralization and sharing data- and causal-knowledge graphs protocols for discovery in federated networks (section 3.3), and *ROBHOOT v.4.0* will integrate *ROBHOOT v.1.0* to *v.3.0* for automated discovery-knowledge graphs in federated cooperation networks.

1.2 Science-to-technology breakthrough that addresses this vision

- Discuss the relevant state-of-the-art and the extent of the advance the project would provide beyond this state-of-the-art
- The state-of-the-art of automated and interpretable discovery is currently a fragmented landscape. The result is a slow response to the rapidly growing global emergency and sustainability challenges. *ROBHOOT* will go beyond the state-of-the-art in automated and interpretable discovery: First, it will introduce automated data- and explainable-knowledge graphs into a more compact and robust discovery-knowledge graph technology. Second, it will fussion the discovery-knowledge graph technology within cooperative federated networks to make discovery a rapidly evolving feature responding to the also rapidly evolving global emergency and sustainability challenges.
- Describe the science-to-technology breakthrough, targeted by the project that would represent the first proof of concept of the envisioned technology.
- Patterns from knowledge-graphs are emerging at a fast pace in specific frontiers, but remains isolated from the discovery process especially in the context of cooperative discovery in federated networks. *ROBHOOT* will go beyond the state-of-the-art of knowledge-graphs by developing data- and causal-knowledge graphs, the discovery-knowledge graphs, in cooperative intelligent federated networks to move knowledge-inspired societies towards reaching global sustainability goals.

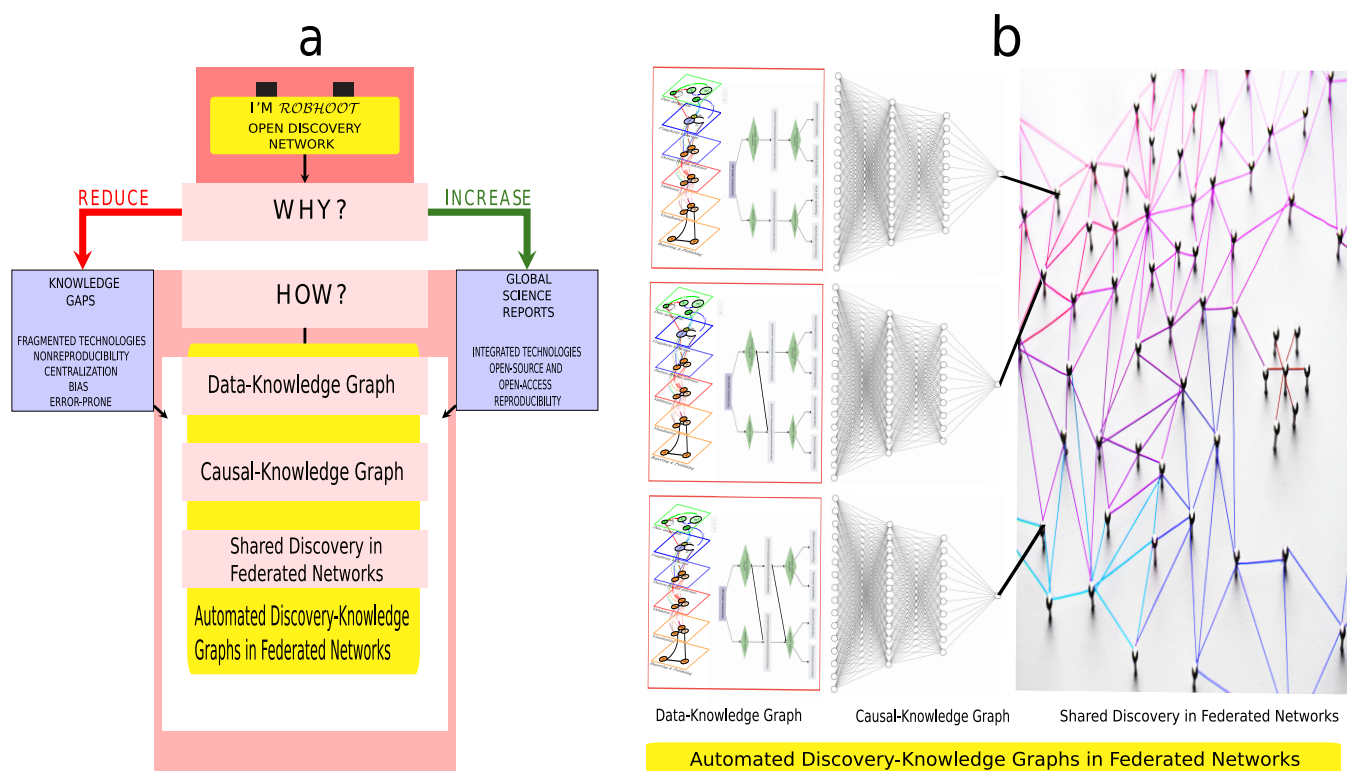


Figure 1: Cooperative Discovery in Federated Networks. *ROBHOOT* integrates data- and causal-knowledge graphs, the discovery-knowledge graph, into cooperative federated networks for a sustainable knowledge-inspired society: **a)** *ROBHOOT* targets global knowledge gaps (red path) and open-access reproducible discovery reports (green path). It integrates three science-enabled technologies: **a,b) Data-Knowledge Graphs** for discovering global data-architecture in emergency and sustainability landscapes. **a,b) Causal-Knowledge Graphs** for fussioning automated and explainable biology-inspired neural networks discovery, and **a,b) Shared Discovery in Federated Networks** for cooperative discovery and forecasting. **Automated Discovery-Knowledge Graphs in Federated Networks** integrates data- and causal-knowledge graphs into intelligent federated networks to generate robust cooperative forecasting to rapidly respond to global emergency and sustainability challenges.

Interconnected global societies constantly face new challenges that need to be rapidly addressed. Yet, technologies integrating data-driven causal inference into intelligent networks providing rapid and global interpretable information when solving complex governance, social, environmental and technological problems are particularly lacking. Depite rapid advances of research platforms for data analytics in the last decade [11, 12, 13, 14, 15, 16, 17, 18, 19], the integration of science-to-technology intelligent automation networks currently lack knowledge-inspired technologies impacting knowledge-inspired societies to help responding to rapidly evolving global sustainability challenges (Figure 1 and Table 1). In this regard, technologies facilitating rapid access to global API-structured data to analyze global data architecture still present challenges. This is particularly relevant in global emergency or sustainability landscapes, where data properties like availability, accuracy and transparency drive constantly emerging feedbacks between questions and scenarios to predict new situations more accurately.

The following are the features of *ROBHOOT*. *ROBHOOT v.1.0* will deploy a data discovery technology to generate question- and data-knowledge graphs for a rapid understanding of global data-architecture. Data-architecture alone is not sufficient to outline predictive scenarios in complex sustainability problems. Therefore, data analytics complementing data-architecture discovery and mechanistic inference is desirable to interpret scenarios in rapidly global emerging challenging situations. In this regard, there are also many gaps in connecting global data-architecture into rapid automated causal-knowledge graphs to facilitate discovery that can be transferred to governance decisions. *ROBHOOT v.2.0* will integrate automated and explainable biology-inspired neural-networks to decipher causal-knowledge graphs from open-ended modeling

| Word | Meaning |
|--------------------------------|--|
| Question-knowledge graph | Technology-driven information extraction from corpus or similar to detect question gaps in multidisciplinary research |
| Data-knowledge graph | Technology-driven information extraction from diverse data-sources to infer global data-architecture |
| Causal-knowledge graph | Technology-driven information extraction to provide explainable/interpretable scenarios on global and complex sustainability challenges |
| Discovery-knowledge multigraph | Novel interactions emerging from the integration of data- and causal-knowledge graphs to provide multidisciplinary responses to global sustainability challenges |
| Automation | Algorithms targeting minimal human-driven interference |
| Knowledge-inspired society | Open-access discovery to take informed decisions in global sustainability challenges |
| Neutral-knowledge generation | Open reproducible reports making transparent the many sources of bias in the discovery process |

Table 1: Glossary of terms.

scenarios. Still, rapidly drawing scenarios from a few labs limit the phase space from where the discovery process is generated. Therefore, the scalability of fully reproducible discovery processes strongly depend on cooperation and learning in decentralized networks. *ROBHOOT v.3.0* will explore sharing protocols in discovery-knowledge graphs in federated networks (section 3.3). Finally, *ROBHOOT v.4.0* will integrate *ROBHOOT v.1.0* to *v.3.0* into automated discovery in federated cooperation networks.

1.3 Interdisciplinarity and non-incrementality of the research proposed

- Describe the research disciplines necessary for achieving the targeted breakthrough of the project and the added value from the interdisciplinarity.
- Explain why the proposed research is non-incremental.

ROBHOOT is a science-enabled multi-feature hybrid-technology for automating interpretable data-driven discovery in intelligent and cooperative federated networks (Figures 1 and 2 and Table 1). It will contain four modules each characterized by a mixture of research disciplines necessary for achieving interdisciplinary breakthrough. *ROBHOOT v.1.0* will be composed by computer scientists and developers targeting API discovery protocols and ETLs algorithms. This module will be complemented with scientists from complex networks to develop quantitative methods for question- and data-knowledge graphs to decipher the existing gaps in data discovery and data-architecture technologies (section 3.1). *ROBHOOT v.2.0* team will be composed by data-scientists trained in deep learning networks and automation algorithms, theoreticians and biologists with expertise in modeling mechanistic and Bayesian networks and biology-inspired neural networks, respectively. The combination of data-scientists, theoreticians and biologists will generate a diverse team targeting synthesis between automated and explainable biology-inspired neural-networks to decipher causal-knowledge graphs from data-architecture properties (section 3.2). *ROBHOOT v.3.0* team will be characterized by combining computer scientists and developers targeting decentralized protocols (federated networks, gnnnet), with social scientist, and scientists specialized in ecology and evolutionary biology. Team for *ROBHOOT v.3.0* will explore cooperation protocols for discovery in federated networks (section 3.3). The complementarity of the teams in the modules one to three will strengthen the collaboration for making *ROBHOOT* a science-enable functional technology in a rapidly evolving digital ecosystem [20]. *ROBHOOT v.4.0* team will combine computer- and data-scientists working in modules one and two, respectively, with developers, biologists, evolutionary biologists and social scientists working in modules 2 and 3. Such a diverse

team will integrate automated and explainable data- and causal-knowledge graphs into federated cooperation networks to generate automated reporting for global emergency and sustainability problems.

ROBHOOT aims to bring global transparency in knowledge generation by acting as an assistant to humans or as a reproducible generator to facilitate sustainability goals of humanity. The multi-feature, science-enabled technology target a reduction in global knowledge gaps while transparently accounting for centralization [4, 7], bias [5], error-prone [6], and non-reproducibility [8] (Figure 1 and Table 1). These features are mostly due to the rapidly evolving digital ecosystem. For example, it is increasing continuously its computing capacity (ref Golem+others), new methods intergating automated and explainable AI are rapidly advancing, and their interconnection to open-source technologies is also rapidly occurring in the digital ecosystem. Yet, targeting automated data- and causal-knowledge graphs in federated cooperative networks still require the filling of many existing technological gaps, from identification and retrieval of heterogeneous data sources, to the integration of explainable modeling and causal inference and the learning capabilities of cooperative forecasting in many interacting automatically evolving agents.

1.4 High risk, plausibility and flexibility of the research approach

- Explain how the research approach relates to the project objectives and how it is suitable to deal with the considerable science-and-technology uncertainties and appropriate for choosing alternative directions and options. (The risks and mitigation plan should be spelled out under the Implementation section).

2 Impact

2.1 Expected impacts

2.2 Measures to maximise impact

3 Implementation

- Describe here the objectives, list of work packages, list of deliverables (Ghentt chart)

Automating the discovery process to tackle rapid global solutions to humanity challenges is highly informative by itself, but a diverse group of scientists across Europe have decided that merely taking discovery alone is not enough. Science is a highly dynamic and global process and there are many paths from where it can be achieved. To understand discovery broadly, these scientists want to advance the automation and cooperative discovery in the global digital ecosystem. To this end, the *ROBHOOT* consortium aims at developing a federated network integrating several technologies into a unified framework. *ROBHOOT* will develop quantitative novel methods such as question-, data-, and causal-knowledge graphs to understand how cooperative discovery networks might help towards knowledge-inspired societies to provide scenarios in face of global sustainability challenges. This strategy is expected to improve early access to discovery to rapidly act in emergency global situations to indentify new emerging targets where automation and global reports can play a key role in knowledge-inspired societies. *ROBHOOT*'s goals will be developed in four different stages with sixteen work packages and milestones and four main deliverables.

3.1 *ROBHOOT* v.1.0: Data Knowledge Graphs

- Rapid API access to build robust and scalable automated interpretable data-driven discovery as an existing need. This is particularly relevant in global emergency or sustainability landscapes, where new questions and scenarios are constantly emerging.
- Data- and Question-Knowledge graphs as solutions for rapid data-driven discovery (defined in Table 1). Which are their features? Which is the state-of-the-art? DKGs will explore similarity patterns of

database to discover existing gaps in data availability and patterns. DKGs will complement QKGs to explore poorly explored questions to new pattern discovery.

- **Global and rapid API access to build robust and scalable question- and data-knowledge graphs as a case study for automated interpretable data-driven discovery.**
- Similar to the idea of having a case study causal-knowledge graph in the *ROBHOOT* v2.0, we can represent a data-knowledge graph following the Robhack examples.

Global and rapid access to data to build robust and scalable question- and data-knowledge graphs is key for automating interpretable data-driven discovery. This is particularly needed in emergency or sustainability challenging situations at the global scale, where new questions and scenarios are constantly emerging. Data access with different privacy requirements, formats, heterogeneity, dimensions, bias and spatiotemporal resolution is the norm and many projects are rapidly emerging to facilitate rapid data access [21, 22, 23, 24]. Yet, available automated science-enabled technologies to build data- and questions-knowledge graphs at the global scale to rapidly inform causal-knowledge graphs about the type of interpretable information that can be extracted, while key to make predictive scenarios, still remain at a very early stage of development despite standard protocols to automate data API access, knowledge extraction, and ETFs algorithms are rapidly advancing [25, 26, 27]. The technologies around automated API data-discovery and interpretable question- and data-knowledge graphs remain difficult to compactly link to the automated causal-knowledge graphs for scientific discovery. *ROBHOOT* v1.0 will deploy the following four work packages each with its own milestone and a final deliverable: data discovery technology (**WP1: API discovery (APID)**) aiming to build question (**WP2: Question-Knowledge graphs (QKGs)**), and data-knowledge graphs (**WP3: Data-Knowledge graphs (DKGs)**), and a case study merging question- and data-knowledge graphs to gain information about global data-architecture (**WP4: Global COVID-19 data-architecture (CODA)**).

3.2 *ROBHOOT* v2.0: Causal Knowledge Graphs

- **Contrasting explainable biologically inspired Causal-Knowledge Graphs for interpretable information when dealing with complex sustainability challenges.**
- **WP5: Causal Spiking Neural Networks (CSN)**
- **Contrasting predictions from Causal Spiking Neural Networks in the framework of Bayesian Space Models to explore open-ended language of models combining Bayesian networks and optimization methods**
- **WP6: Bayesian Space Models (BSM)**
- **Case study**
- **Covid-19 Causal Knowledge Graph**

AI is rapidly advancing in automated discovery (i.e., AutoML [10]) making more transparent the processes underlying the discovery (i.e., Explainable or interpretable AI [28, 29]). Yet, automated and explainable discovery methods are still at an incipient stage of integration, particularly in open-ended Bayesian machines [14]. This is particularly relevant in the context of biology, brain research, and evolutionary biology techniques where making automatic interpretation of complex systems can provide scenarios to help disentangling complex sustainability problems for humanity. *ROBHOOT* v2.0 will develop novel causal knowledge graphs integrating automated and explainable discovery accounting for biologically inspired neural networks and evolutionary biology techniques [30, 31]. Automatic interpretation of the causal processes underlying empirical patterns will be explored in a series of neuromorphic computing scenarios using neural networks of evolving spiking neurons (**WP5: Causal Spiking Networks (CSN)**) (Figure 2).

Causal Spiking Neurons will explore open-ended language of models with varying biologically relevant functions like code insertions, deletions, inversions and other molecular and genotype-phenotype processes to search for automated biologically inspired solutions to complex empirical patterns (**WP6: Causal-Knowledge Graphs (CKG)**, Figure 3a). Causal knowledge graphs will help to enhance the connection between automated

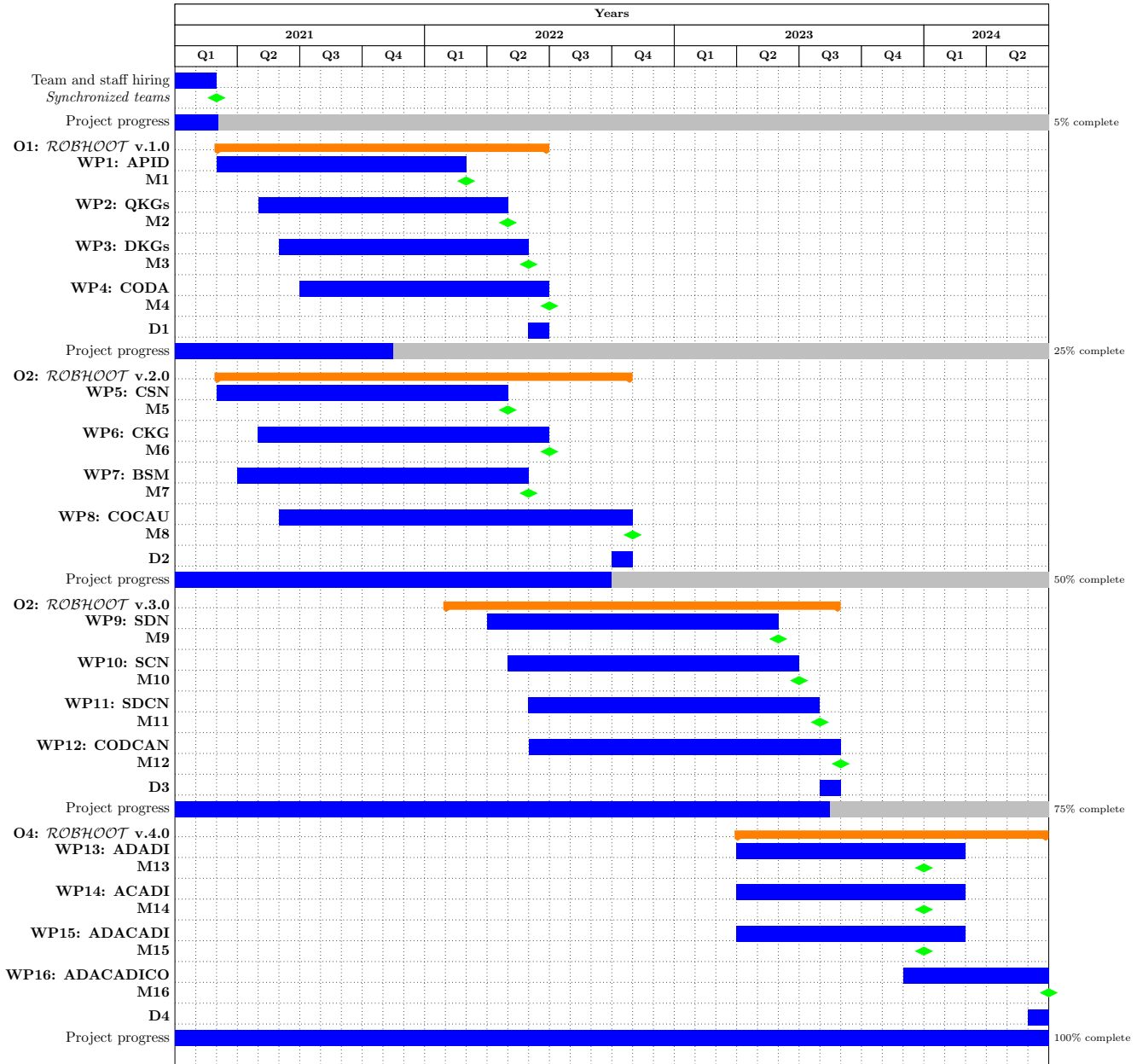


Figure 2: Roadmap: *ROBHOOT v.1.0* packages and milestones one to four bring question- and data-knowledge graphs into a global data-architecture map (Deliverable, **D1**). *ROBHOOT v.2.0* packages and milestones five to eight fusion automation and interpretable patterns into causal-knowledge graphs (Deliverable, **D2**). *ROBHOOT v.3.0* packages and milestones nine to twelve connect shared discovery to federated networks (Deliverable, **D3**), and *ROBHOOT v.4.0* packages and milestones thirteen to sixteen bring automated discovery-knowledge graphs in federated networks to close knowledge-gaps in the rapidly emerging global sustainability challenges (Deliverable, **D4**).

and explainable AI throughout prediction and knowledge power (Figure 3b). Automated and interpretable data inference still present many challenges in the context of multidimensional landscapes [32, 33]. This is particularly relevant in Earth and Ecosystem science and Biodiversity and Sustainability science [34], where merging automation to interpretable data will increase human ability to make stronger inferences about future sustainability challenges and solutions. In order to make inference from complex data more robust we will contrast predictions from Causal Spiking Networks in the framework of Bayesian Space Models to explore open-ended language of models combining Bayesian networks and optimization methods (**WP7: Bayesian Space Models (BSM)**). The Bayesian space models module will ensure the search, the evaluation of models,

trading-off complexity, fitting to the data and quantify resource usage [12, 14]. *ROBHOOT* v2.0 will deploy the COVID-19 pandemic as a case study to automatically infer locally interpretable causal-knowledge graphs at global scale (WP8: COVID-19 Causal-Knowledge Graphs (COCAU)).

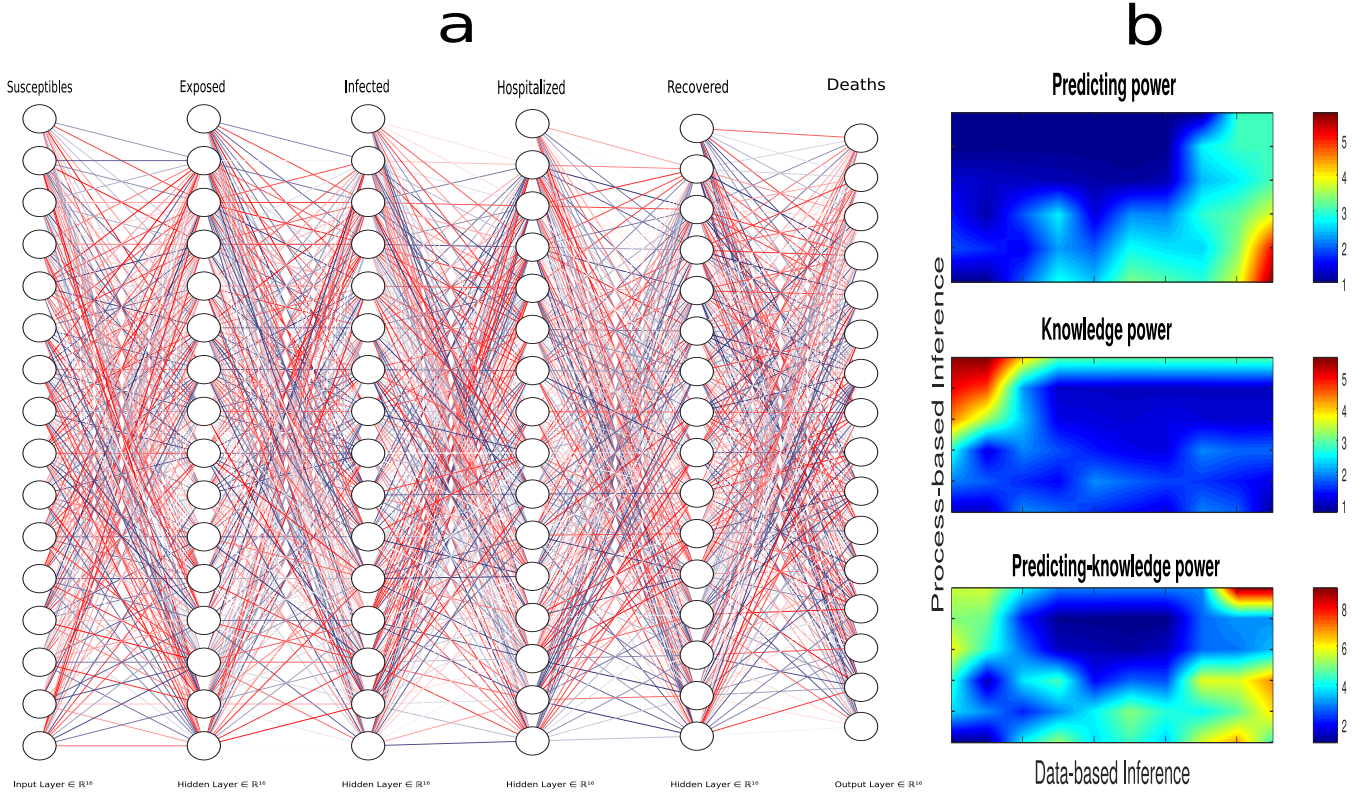


Figure 3: Causal-Knowledge Graphs. a) They contain features of deep learning networks and process-based modeling to make empirical patterns interpretable. For epidemiology models for example, all input, hidden and output layers can be “converted” to age-class susceptibles (input), exposed, infected, hospitalized, and recovered (all hidden), and deaths (output). The same applies to edges connecting each pair of nodes. Therefore, a Causal-Knowledge Graph recovers the temporal sequence of events that best predict the output layer (i.e., Deaths). b) Causal-Knowledge Graphs aim to integrate predicting and knowledge power into interpretable knowledge (bottom, b). x-axis represents “Data-based inference” (i.e., gradient of AI methods from low (left) to high (right) predicting power). y-axis represents “Process-based inference” (i.e., gradient of process-based methods from low (bottom left) to high (top left) knowledge power). The gradient of predicting power map (top) shows a hot spot red area in the bottom right highlighting the region where AI methods best predict the empirical data. The gradient of knowledge power map (middle) shows a red hot spot in the top left highlighting the region where the best mechanistic understanding occur. The predicting-knowledge power map (bottom) shows the sum of the two previous maps highlighting a red hot spot where predicting and knowledge power occur. Red and blue lines show important and non-important roles in the predicting-knowledge power map, respectively.

3.3 *ROBHOOT* v3.0: Shared Discovery in Federated Networks

- A science-based explainable-knowledge graph technology is not enough if we aim to globally contrast robustly paths for interpretation of causal process predicting global data-architectures.
- A science-based technology to develop shared cooperative forecasting aiming to find robust populations of results in face of rapidly emerging global emergency and sustainability challenges.

A science-based explainable technology is not enough if we aim to globally contrast robustly paths of

interpretable causal processes predicting the global data-architecture obtained from the data-and causal-knowledge graphs. In this regard, sharing data- and causal-knowledge graphs are needed to develop scalable cooperative forecasting and strong inference. *ROBHOOT* v.3.0 will focus on the development of protocols in digital networks to embed explainable discovery-knowledge graphs into global cooperation schemes to increase robustness and reproducibility for rapid reporting generation [3]. Technologies in decentralized digital ecosystems are rapidly advancing in a variety of sectors. Most progress is coming in the scalability, security and decentralization trade-offs [3, 35, 36, 37, 38, 39]. In the science ecosystem, only a few applications of open decentralized technologies exist [7]. Yet, sharing reproducible data- and causal-knowledge graphs, the discovery-knowledge graphs, along networks to facilitate cooperative forecasting is currently not in place.

ROBHOOT v.3.0 will deploy the following four work packages and milestone and a final deliverable for sharing discovery in networks: sharing question- and data-knowledge graphs, the data-architecture, in federated networks using reproducible-knowledge graphs. **WP9 to WP11, Sharing data-architecture (SDN), causal-knowledge graphs, and discovery-knowledge graphs in federated networks (SCN), respectively:** These packages will explore scalability and security properties of sharing knowledge graphs using different topologies of federated networks. For example, there can be data-knowledge graphs belonging to specific categories of data but they might be integrated with different data types to gain global understanding of data-architecture (i.e., like letting people on different social networks follow each other). Despite the dramatic rise in global pandemics during the last decade (i.e., the SARS pandemic in 2003, to Avian Influenza in 2006, H1N1 in 2009, Ebola in 2014, the appearance of the Zika virus in Latin America in 2015, and the current Covid-19 pandemic), with these developments inextricably bound up in modern socio-technical developments and processes of globalization, science-based technologies facilitating rapid sharing of information to mitigate risks and enhance response efficiency with globally informed scenarios are particularly lacking [2]. The last work package of *ROBHOOT* v.3.0, **(WP12: Sharing discovery-knowledge graphs in federated networks, COVID-19 as a case study (CODCAN))**, will deploy sharing data-architecture and causal-knowledge graphs, the discovery-knowledge graphs, in federated networks for the COVID-19 case study.

3.4 *ROBHOOT* v.4.0:

Automated Discovery-Knowledge Graphs in Federated Networks

- Science-based explainable discovery-knowledge graphs shared in a federated network is not enough if most contributions can not be reproduced and strongly depend on competitive schemes. This might be particularly relevant in the face of rapidly responding to global emergency and sustainability challenges.
- A science-based technology to develop automatically shared cooperative forecasting aiming to find robust populations of results in the face of rapidly emerging global emergency and sustainability challenges.

Science-based explainable discovery-knowledge graphs shared in a federated network is not enough if most contributions can not be reproduced and strongly depend on competitive schemes. This is particularly relevant in the face of rapidly responding to global emergency and sustainability challenges. For example, in the context of the COVID-19 pandemic, and in many other situations, models and predictions abound, but no one can say with certainty what the course of the virus will be, and much less the impact the pandemic will have on people and societies [40]. In such an scenario, making shared contributions fully reproducible and containing a mixture of human and automatically generated discovery-knowledge graphs can guide us to improve the robustness of the forecasting. In this regard, shared data- and causal-knowledge graphs, the discovery-knowledge graphs, containing a mixture of human and automatically generated discovery-knowledge graphs, from where cooperative forecasting can estimate robust populations, might help to rapidly generate global reports of future scenarios in real-time. *ROBHOOT* v.4.0 will connect interpretable discovery-knowledge graphs to automation to study the properties of scalable cooperative forecasting and strong inference in the face of global sustainability challenges.

ROBHOOT v.4.0 will deploy the following four work packages and milestone and a final deliverable for automated discovery in federated networks: **WP13 to WP15, Automated data- (ADAN), causal- (ACAN), and discovery-knowledge graphs (ADIN) in federated networks, respectively** will deploy and share data-,

causal- and discovery-knowledge graphs from automatically and cooperatively generated knowledge graphs. The last work package of *ROBHOOT* v.4.0, **WP16: Automated Cooperative Forecasting of Discovery-Knowledge Graphs for the COVID-19 case study (COADIN)**, will deploy and share automatically generated discovery-knowledge graphs in federated networks for the COVID-19 case study to contrast cooperative forecasting under populations of scenarios.

3.5 Research methodology and work plan – Work packages, deliverables

3.6 Management structure, milestones and procedures

- Describe the organisational structure and the decision-making (including a list of milestones (table 3.2a))
- Explain why the organisational structure and decision-making mechanisms are appropriate to the complexity and scale of the project.
- Describe any critical risks, relating to project implementation, that the stated project's objectives may not be achieved. Detail any risk mitigation measures. Please provide a table with critical risks identified and mitigating actions (table 3.2b) and relate these to the milestones.

3.7 Consortium as a whole

- The individual members of the consortium are described in a separate section 4
- Describe the consortium. Explain how it will support achieving the project objectives. Does the consortium provide all the necessary expertise? Is the interdisciplinarity in the breakthrough idea reflected in the expertise of the consortium?
- In what way does each of the partners contribute to the project? Show that each has a valid role and adequate resources in the project to fulfil that role. How do the members complement one another? Other countries and international organisations: If one or more of the participants requesting EU funding is based in a country or is an international organisation that is not automatically eligible for such funding (entities from Member States of the EU, from Associated Countries and from one of the countries in the exhaustive list included in General Annex A of the work programme are automatically eligible for EU funding), explain why the participation of the entity in question is considered essential for carrying out the action on the grounds that participation by the applicant has clear benefits for the consortium.

ROBHOOT is a science-enabled multi-feature hybrid-technology. Multi-feature and hybrid imply, from one side, algorithms and quantitative metrics to bring question- and data-knowledge graphs, the data-discovery process, to the global data-architecture picture (*ROBHOOT* v.1.0, section 3.1). This pattern-driven exercise requires a mixture of researchers, from computer scientists and developers to researchers working in complex networks from the quantitative and epistemological angles. From the other side, fusing automation and interpretable patterns along causal-knowledge graphs add complementarity expertise to the global data-architecture picture. Now the skills focus on data-scientists trained in deep learning (interpretable) networks and automation algorithms, theoreticians with expertise in Bayesian inference, and evolutionary biologists and neurobiologists with expertise in empirical patterns and biology-inspired neural modeling networks (*ROBHOOT* v.2.0, section 3.2). Despite modules from *ROBHOOT* v.1.0 and *ROBHOOT* v.2.0 will produce specific milestones and deliverables (Figure 2), they will also recombine data-knowledge graphs into causal-knowledge graphs, the discovery-knowledge graphs, to build an interdisciplinarity-hybrid science-enabled technology that can be compactly converted into user-friendly open-software. Therefore, interdisciplinarity enters not only at the intra-module development stage, but also at the inter-module stage where functionality driven discovery-knowledge graphs can trigger a breakthrough that is reflected in the complementarity skills of the consortium (section 4.1). The first two modules from *ROBHOOT* contain researchers from four countries, Estonia, Spain, Switzerland and Sweden, with expertise in Deep Learning networks and AI, Complex networks (quantitative metrics, epistemology and modeling), Theory

with Bayesian inference expertise and Modeling in Evolutionary biology and Biology and Computer software design, respectively.

The *ROBHOOT* consortium wants to understand discovery broadly. We want to take advantage and advance the rapidly evolving digital ecosystem by making cooperative discovery a fundamental feature of it. For this purpose, a science-based automated and interpretable technology is not enough if we aim to globally contrast robustly paths for robust interpretation of causal processes for delineating plausible scenarios in the face of global sustainability challenges. To achieve scalability for the discovery-knowledge graphs, sharing will be the excellency feature of *ROBHOOT v.3.0* (section 3.3). It will develop protocols for sharing data- and causal-knowledge graphs, the discovery-knowledge graphs, for making them reproducible in the context of cooperative federated networks. The team forming *ROBHOOT v.3.0* will therefore require quite a lot of contrasting skills. First, developers working in decentralized and security protocols. Second, social scientists in collaboration to developers aiming to build user-friendly open-access interfaces to explore alternative scenarios of sharing. Third, data-scientists working in reproducible-knowledge graphs to make the discovery-knowledge graphs “reproducible shared objects”. Notice, a strong complementarity but also a functionality driven cooperation between (*ROBHOOT v.1.0*, *v.2.0* and *v.3.0*) is required to make the discovery-knowledge graphs reproducible and functionally accessible in a cooperative federated network. The last module of (*ROBHOOT*) adds the automated feature of discovery-knowledge graphs in federated networks for making cooperative forecasting a synchronized and standard global property in the face of emergency or sustainability challenging situations. The last two modules from *ROBHOOT* contain researchers from XX countries, XX, XX, Spain, Switzerland and XX, with expertise in protocols in decentralized algorithms, security in computing networks and distributed networks, automation algorithms and social networks

3.8 Resources to be committed

- Please make sure the information in this section matches the costs as stated in the budget table in section 3 of the administrative proposal forms, and the number of person months, shown in the detailed work package descriptions. Please provide the following:
- a table showing number of person months required (table 3.4a)
- a table showing ‘other direct costs’ (table 3.4b) for participants where those costs exceed 15% of the personnel costs (according to the budget table in section 3 of the administrative proposal forms)

4 Members of the consortium

4.1 Participants (applicants)

- For each participant, provide the following: a description of the legal entity and its main tasks, with an explanation of how its profile matches the tasks in the proposal
- a curriculum vitae or description of the profile of the persons, including their gender, who will be primarily responsible for carrying out the proposed research and/or innovation activities. Indicate each person who would be a first-time participant to FET under Horizon 2020
- a list of up to 5 relevant publications, and/or products, services (including widely-used datasets or software), or other achievements relevant to the call content
- redList of up to 5 relevant previous projects or activities, connected to the subject of this proposal
- a description of any significant infrastructure and/or any major items of technical equipment, relevant to the proposed work
- if operational capacity cannot be demonstrated at the time of submitting the proposal, describe the concrete measures that will be taken to obtain it by the time of the implementation of the task.1

4.2 Third parties involved in the project (including use of third party resources)

- For each participant, does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted) Y/N If yes, please describe and justify the tasks to be subcontracted
- Does the participant envisage that part of its work is performed by linked third parties² Y/N If yes, please describe the third party, the link of the participant to the third party, and describe and justify the foreseen tasks to be performed by the third party
- Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement) Y/N If yes, please describe the third party and their contributions
- Does the participant envisage that part of the work is performed by International Partners³ (Article 14a of the General Model Grant Agreement)? Y/N If yes, please describe the International Partner(s) and their contributions.

5 Ethics and Security

This section is not covered by the page limit.

5.1 Ethics

For more guidance, see the document "How to complete your ethics self-assessment". If you have entered any ethics issues in the ethical issue table in the administrative proposal forms, you must:

- submit an ethics self-assessment, which:
- describes how the proposal meets the national legal and ethical requirements of the country or countries where the tasks raising ethical issues are to be carried out;
- explains in detail how you intend to address the issues in the ethical issues table, in particular as regards: research objectives (e.g. study of vulnerable populations, dual use, etc.) research methodology (e.g. clinical trials, involvement of children and related consent procedures, protection of any data collected, etc.)
- the potential impact of the research (e.g. dual use issues, environmental damage, stigmatisation of particular social groups, political or financial retaliation, benefit-sharing, misuse, etc.)
- If you plan to request these documents specifically for the project you are proposing, your request must contain an explicit reference to the project title.

5.2 Security

- activities or results raising security issues: (YES/NO)
- EU-classified information as background or results: (YES/NO)

References

- [1] Matías E Mastrángelo, Natalia Pérez-Harguindeguy, Lucas Enrico, Elena Bennett, Sandra Lavorel, Graeme S Cumming, Dilini Abeygunawardane, Leonardo D Amarilla, Benjamin Burkhard, Benis N Egoh, Luke Frishkoff, Leonardo Galetto, Sibyl Huber, Daniel S Karp, Alison Ke, Esteban Kowaljow, Angela Kronenburg-García, Bruno Locatelli, Berta Martín-López, Patrick Meyfroidt, Tuyeni H Mwampamba, Jeanne Nel, Kimberly A Nicholas, Charles Nicholson, Elisa Oteros-Rozas, Sebataolo J Rahlao, Ciara Raudsepp-Hearne, Taylor Ricketts, Uttam B Shrestha, Carolina Torres, Klara J Winkler, and Kim Zoeller. Key knowledge gaps to achieve global sustainability goals. *Nature Sustainability*, 2(12):1115–1121, 2019. ISSN 2398-9629. doi: 10.1038/s41893-019-0412-1. URL <https://doi.org/10.1038/s41893-019-0412-1>.
- [2] Christopher Wilson and Maria Gabrielsen Jumbert. The new informatics of pandemic response: humanitarian technology, efficiency, and the subtle retreat of national agency. *Journal of International Humanitarian Action*, 3(1), 2018. ISSN 2364-3412. doi: 10.1186/s41018-018-0036-5.
- [3] Johnny Dilley, Andrew Poelstra, Jonathan Wilkins, Marta Piekarska, Ben Gorlick, and Mark Friedenbach. Strong Federations: An Interoperable Blockchain Solution to Centralized Third-Party Risks. 2016. URL <http://arxiv.org/abs/1612.05491>.
- [4] H. Inhaber. Changes in centralization of science. *Research Policy*, 6(2):178–193, apr 1977. ISSN 0048-7333. doi: 10.1016/0048-7333(77)90024-5. URL <https://www.sciencedirect.com/science/article/abs/pii/0048733377900245>.
- [5] John P a Ioannidis. Why most published research findings are false. *PLoS medicine*, 2(8):e124, aug 2005. ISSN 1549-1676. doi: 10.1371/journal.pmed.0020124. URL <http://www.ncbi.nlm.nih.gov/pubmed/16060722>.
- [6] Ferric C Fang and Arturo Casadevall. Retracted Science and the Retraction Index. *Infection and Immunity*, 79(10):3855 LP – 3859, oct 2011. doi: 10.1128/IAI.05661-11. URL <http://iai.asm.org/content/79/10/3855.abstract>.
- [7] Vlad Günther and Alexandru Chirita. " Scienceroot " Whitepaper. 2018. URL <https://www.scienceroot.com/>.
- [8] Tom E. Hardwicke, Maya B. Mathur, Kyle MacDonald, Gustav Nilsson, George C. Banks, Mallory C. Kidwell, Alicia Hofelich Mohr, Elizabeth Clayton, Erica J. Yoon, Michael Henry Tessler, Richie L. Lenne, Sara Altman, Bria Long, and Michael C. Frank. Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science*, 5(8):180448, sep 2018. ISSN 20545703. doi: 10.1098/rsos.180448. URL <https://doi.org/10.1098/rsos.180448>.
- [9] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A Survey on Bias and Fairness in Machine Learning. 2019. URL <http://arxiv.org/abs/1908.09635>.
- [10] Esteban Real, Chen Liang, David R. So, and Quoc V. Le. AutoML-Zero: Evolving Machine Learning Algorithms From Scratch. 2020. URL <http://arxiv.org/abs/2003.03384>.
- [11] Melnik S., Gubarev A., Long J. J., Romer G., Shivakumar S., Tolton M., and T Vassilakis. Dremel: Interactive Analysis of Web-Scale Datasets. *Proc. of the 36th International Conference on Very Large Data Bases (VLDB)*, pages 159–176, 2010.
- [12] Christian Steinruecken, Emma Smith, David Janz, James Lloyd, and Zoubin Ghahramani. The Automatic Statistician. pages 161–173. 2019. ISBN 9783030053185. doi: 10.1007/978-3-030-05318-5_9.

-
- [13] *Modulos*. URL <http://www.modulos.ai/>.
- [14] Roger Guimerà, Ignasi Reichardt, Antoni Aguilar-Mogas, Francesco A. Massucci, Manuel Miranda, Jordi Pallarès, and Marta Sales-Pardo. A Bayesian machine scientist to aid in the solution of challenging scientific problems. *Science Advances*, 6(5):eaav6971, 2020. ISSN 23752548. doi: 10.1126/sciadv.aav6971.
- [15] *Google AI*. URL <https://ai.google/>.
- [16] *Iris*. URL <https://iris.ai>.
- [17] *EaseML*. URL <https://github.com/DS3Lab/easeml>.
- [18] *Datarobot*. URL <https://www.datarobot.com/>.
- [19] *Aito*. URL <https://aito.ai/>.
- [20] Cesar Soto-Valero, Amine Benelallam, Nicolas Harrand, Olivier Barais, and Benoit Baudry. The emergence of software diversity in maven central. *IEEE International Working Conference on Mining Software Repositories*, 2019-May:333–343, 2019. ISSN 21601860. doi: 10.1109/MSR.2019.00059.
- [21] *Openstreet*, . URL <https://www.openstreetmap.org/{#}map=5/51.509/-0.110>.
- [22] *Blue-cloud*. URL <https://www.blue-cloud.org/>.
- [23] *HOT*. URL <https://www.hotosm.org/>.
- [24] *Elixir open data*. URL <https://elixir-europe.org/>.
- [25] J. Fan, A. Kalyanpur, D. C. Gondek, and D. A. Ferrucci. Automatic knowledge extraction from documents. *IBM Journal of Research and Development*, 56(3.4):5:1–5:10, may 2012. ISSN 0018-8646. doi: 10.1147/JRD.2012.2186519. URL <http://ieeexplore.ieee.org/document/6177723/>.
- [26] *APIs*. URL <https://apis.guru/openapi-directory/>.
- [27] *Open Knowledge Foundation*, . URL <https://discuss.okfn.org/>.
- [28] Yolanda Gil and Bart Selman. A 20-Year Community Roadmap for Artificial Intelligence Research in the US. aug 2019. URL <https://bit.ly/2ZNVBvh><http://arxiv.org/abs/1908.02624>.
- [29] Giuseppe Futia and Antonio Vetrò. On the integration of knowledge graphs into deep learning models for a more comprehensible AI-Three challenges for future research. *Information (Switzerland)*, 11(2), 2020. ISSN 20782489. doi: 10.3390/info11020122.
- [30] Wolfgang Maass. Noise as a resource for computation and learning in networks of spiking neurons. *Proceedings of the IEEE*, 102(5):860–880, 2014. ISSN 00189219. doi: 10.1109/JPROC.2014.2310593.
- [31] Wolfgang Maass. To Spike or Not to Spike: That Is the Question. *Proceedings of the IEEE*, 103(12):2219–2224, 2015. ISSN 00189219. doi: 10.1109/JPROC.2015.2496679.
- [32] Anthony O’Hare. Inference in High-Dimensional Parameter Space. *Journal of Computational Biology*, 22(11):997–1004, 2015. ISSN 1066-5277. doi: 10.1089/cmb.2015.0086. URL <http://online.liebertpub.com/doi/10.1089/cmb.2015.0086>.
- [33] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. pages 1–10, 2019. URL <http://arxiv.org/abs/1911.01429>.

-
- [34] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and & Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature*. ISSN 0028-0836. doi: 10.1038/s41586-019-0912-1. URL www.nature.com/nature.
- [35] Golem. The Golem Project Crowdfunding Whitepaper. *Golem.Network*, (November):1–28, 2016. URL <https://golem.network/crowdfunding/Golemwhitepaper.pdf>.
- [36] Nikolai Durov. Telegram Open Network. pages 1–132, 2017.
- [37] Elli Androulaki, Artem Barger, Vita Bortnikov, Srinivasan Muralidharan, Christian Cachin, Konstantinos Christidis, Angelo De Caro, David Enyeart, Chet Murthy, Christopher Ferris, Gennady Laventman, Yacov Manevich, Binh Nguyen, Manish Sethi, Gari Singh, Keith Smith, Alessandro Sorniotti, Chrysoula Stathakopoulou, Marko Vukolić, Sharon Weed Cocco, and Jason Yellick. Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains. *Proceedings of the 13th EuroSys Conference, EuroSys 2018*, 2018-Janua, 2018. doi: 10.1145/3190508.3190538.
- [38] Ocean Protocol Foundation, BigchainDB GmbH, and DEX Pte. Ltd. Ocean Protocol: A Decentralized Substrate for AI Data & Services Technical Whitepaper. pages 1–51, 2018. URL <https://oceanprotocol.com/>.
- [39] BigchainDB GmbH. BigchainDB: The blockchain database. *BigchainDB. The blockchain database.*, (May):1–14, 2018. doi: 10.1111/j.1365-2958.2006.05434.x. URL <https://www.bigchaindb.com/whitepaper/bigchaindb-whitepaper.pdf>.
- [40] Remarkable times call for remarkable collaboration . (April), 2020.