ROBHOOT
Automated Discovery in Global Federated Networks
v.1.0

April 6, 2020

Summary

Global sustainability is a major goal of humanity. Many studies have shown global sustainability could be achieved by strengthening transparency and feedbacks between social, ecological, technological and governance systems. Sustainability goals, however, strongly depend on global access to evidence-, research-, and discovery-based knowledge gaps. Yet, the science ecosystem lacks open-source technologies narrowing down global knowledge gaps. We introduce a rapid automated discovery technology targeting knowledge gaps throughout reproducible learning in federated networks. A federated discovery network encompasses a hybrid-automated-technology to lay out the foundation of an open-, cooperative-science ecosystem to automate discovery strengthening rapid solutions and robustness at the global scale in emergency or sustainability problems. The project summarized here is not set out to deliver a finished rapid automated discovery network, but to provide the architecture of a science-enabled technology as a proof-of-principle to connect global human problems to neutral-knowledge generation in 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
- Causal knowledge graphs as science-enabled technology facilitating process based governance and data interpretation
- Describe how this vision surpasses substantially any technological paradigms that currently exist or are under development.
- Technological paradigms for rapid discovery are currently based on competitive schemes instead of learning and cooperation among many labs in the context of a federated network. Technological paradigm is also highly fragmented, partly solve reproducibility, and are mostly developed in close-source software.
- 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).

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 intelligent machines, open discovery-based technologies facilitating global access to informed decisions when solving complex social, environmental and technological problems are particularly lacking. How can discovery driven intelligent machines help to reach global sustainability goals by reducing knowledge gaps? The *ROBHOOT* project introduces new concepts to global knowledge gaps. Current technologies for automated scientific inquiry are highly fragmented, partly solve reproducibility and are mostly developed in close-source software. Thus, despite the importance of global access to knowledge gaps for reaching sustainability goals, open-source technologies facilitating learning, full research cycle reproducibility, and cooperation in federated networks are lacking. The goal of *ROBHOOT* is to propose a new hybrid-technology concept integrating question-, data-, reproducibility-, and causal-based knowledge-graphs to lay the foundation for a novel scientific discovery technology.

More than half (3.9 billion) of the global population is now online and using the Internet,[2]which represents a more inclusive global information society. And 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.

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: Automated causal inference exploring cuting-edge neural networks
- Describe the science-to-technology breakthrough, targeted by the project that would represent the first proof of concept of the envisioned technology: Collapsing evidence- and research-based knowledge gaps for a sustainable knowledge-inspired society

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 aims to be a hybrid-technology accounting for many features (Tables 1 to 3). Producing such a multi-feature technology requires multidisciplinarity teams making contributions for each of the Robhoot features while integrating all these features in a rapidly evolving digital ecosystem. In this regard, the project aims to put together data and computer scientists (i.e., distributed computing, open-source software development), scientists from the physics of complex systems (i.e., multilayer networks), artificial intelligence (i.e., deep learning and automation) and the biology, ecology and evolution of social, natural and technological ecosystems.

Technologies with the capacity to compactly account for neutral, borderless, immutable, and open-access information in hybrid, trusted-untrusted peer-to-peer interactions, accounting for the multilayer nature of science and engineering are currently not in place. Producing such a technology will require integrating expertise from disparate disciplines like multilayer networks, deep learning, automation algorithmics, and distributed technologies. The integration of these disciplines will require to go beyond domain boundaries. Specifically, we will merge scientists and engineers from data and computer science, the physics of complex systems, artificial intelligence and the biology, ecology and evolution of social, natural and technological ecosystems to develop a "de novo" technology: the synthesis of automated knowledge generation in a neutral, borderless and

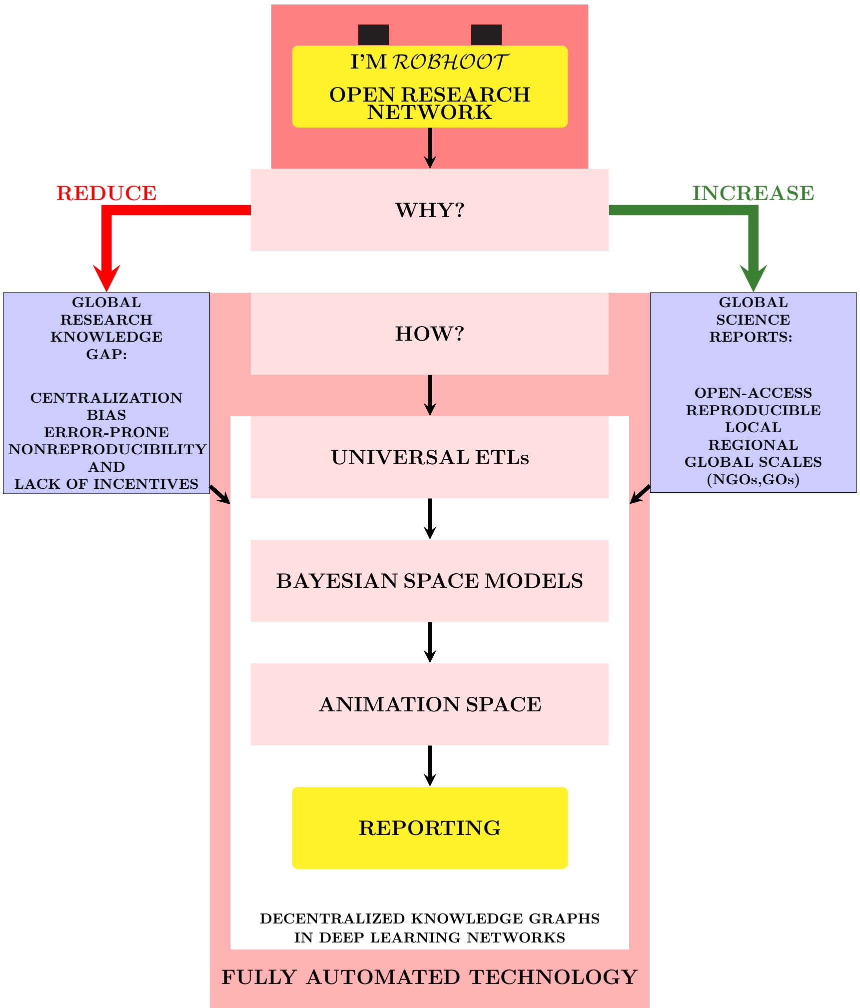

Figure 1: Discovery network technology. *ROBHOOT targets global knowledge gaps (red path) and open-access fully reproducible discovery reports (green path). ROBHOOT is open-source science-based automated technology decentralizing reproducible knowledge graphs. ROBHOOT integrates 1) question-based knowledge gaps with data extraction, transformation and loading algorithms for database integration and complexity reduction (***qETLs***), 2)* **Bayesian Space Models** *accounting for open-ended model optimization techniques, 3)* **Animation Space** *visualizing fitting procedures and pattern generation connecting empirical data and open-ended models, and 4) open and globally accessible* **Reporting** *formalized in natural processing language.*

immutable network synthesized anew from existing open-source projects like Renku, Fabric and gitchain.

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

- Each of the expected impacts listed in the work programme:
- Scientific and technological contributions to the foundation of a new future technology.
- Potential for future social or economic impact or market creation.
- Building leading research and innovation capacity across Europe by involvement of key actors that can make a difference in the future, for example excellent young researchers, ambitious high-tech SMEs or first-time participants 2 to FET under Horizon 2020
- any substantial impacts not mentioned in the work programme, that would enhance innovation capacity; create new market opportunities, strengthen competitiveness and growth of companies, address issues related to climate change or the environment, or bring other important benefits for society.

The following are the general and the specific impacts according to our objectives, working packages and deliverables:

- Automated knowledge-based network technology

  The integration between open-source data integration and inference schemes, the interlayer automation (O1: Multilayer), will allow for the systematic exploration of robust knowledge-based patterns when exploring the population of KGs. This is in sharp contrast to existing AI technologies mostly oriented to prediction without knowledge-based understanding (refs). Despite open-source ETLs are rapidly evolving towards accounting for many aspects of data integration (formats, historical-real time, storage, dimensions, size, bias and spatiotemporal resolution), there is a missing component in quantifying the robustness of knowledge that integrated data can provide. Automated populations of KGs connecting cutting-edge open-source ETLs to inference classification schemes can provide the quantification of robustness in knowledge-based patters for future predictive technologies.
- Open immutable knowledge in untrusted digital peer-to-peer ecosystems

  The open access of immutable accumulation of knowledge in untrusted digital peer-to-peer ecosystems: Social, environmental and economic impact to facilitate global access to transparent knowledge. ETLs are rapidly evolving towards accounting for many key aspects of data integration: Data manipulation across formats (CloverDX), merging of historical and real-time streaming data pipelines (i.e., Kafka) and data structures facilitating the storage and access of large amounts of data (i.e., clickhouse.) Our research methodology will be focused on developing an automated workflow using the geographically distributed cloud on computing and storage to test the robustness of data integration metrics across gradients of simulated data containing dimensions, biases, sizes, formats, temporal and spatial resolution (should we be more domain specific here? Or should we stay general and thinking broadly about simulated data with along complexity gradients and explore data integration metrics? How is the SDSC dealing with data integration for Renku? We anticipate implementation of an automated end-to-end research cycle within an open ledger to facilitate real-time open-access neutral data-rule-knowledge to gain informed decisions to help solve complex social, environmental and technological problems. This facilitation might occur for local, regional and global problems in many fronts. Specifically, open deep ledger knowledge networks might have an impact in the following five areas

- The identification of gaps in research paths not explored consequence of lack of synthesis in interdisciplinary research: The creation of new markets opportunities obtained from exploring these gaps and the development of comparative method in the science of science and citizen data science.
- The merging of prediction and explanatory power in open science to gain synergy between AI open predictive tools and ruled-based pattern inference creating a more balanced pattern and process inference interaction. Recent examples of AI algorithms playing chess and go using brute force deep learning models or rule-based algorithms have discovered the power...: The integration between prediction and understanding power to facilitate explanatory synthesis.
- The automation of reproducible open knowledge will facilitate the reusability, repeatability, and replicability of research outputs. The open access knowledge for governance transparency.

2.2 Measures to maximise impact

- Dissemination and exploitation of results
  1. G4 will launch a testnet to help disseminate the main results of the deep ledger knowledge network. The launch will have invited NGO's and GO across disciplines and social, economical and technological sectors.
  2. The Robhoot Open network will be launched as a Biodiversity research network to integrate the existing public databases and crowdsource data collections into the automated KGs and ledger network to facilitate NGOs, GO and other organizations transparency and governance in Biodiversity management.
  3. The project aims to publish its main findings in top open scientific journals to communicate the global impact of a deep ledger knowledge network for transparency and governance across social and economical sectors.
- Communication activities
  1. The contribution in communication of the Swiss Data Science Center, Switzerland
  2. Contribution of the Wyss center
  3. Contribution of Ifisc, Spain

3 Implementation

- Describe here the objectives, list of work packages, list of deliverables (mention 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 how discovery broadly, these scientists want to share the advantages of cooperative discovery in the global digital ecosystem. To this end, they formed the *ROBHOOT* consortium with the goal of developing a federated network integrating several technologies into a unified framework. *ROBHOOT* scientists 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 anticipate global sustainability challenges. This strategy is expected to improve early access to discovery to rapidly act in emergency global situations, and 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.

3.1 *ROBHOOT* v.1.0: Data Knowledge Graphs

Conceptual and technical needs for DKGs

- **API Discovery: APIDIS** Standard protocols advancing to build API database to facilitate data discovery (API projects). However, automated API discovery to decentralized API database can facilitate robustness in the medium and long term. APIDIS will automate existing open-source data extraction, transformation and load with unique features (i.e., formats, historical-real time, storage, dimensions, size, sampling bias and spatiotemporal resolution, etc) (Figures 1 and 2a, two top layers).
- **Question Knowledge graphs: QKGs** QKGs will be explored to detect poorly explored questions but important for multidisiplinarity in the discovery process. Finds the gaps of question knowledge graphs to discover relevant transdisciplinary topics.
- **Data Knowledge graphs: DKGs** DKGs will explore similarity patterns of database similarity to discover existing gaps in empirical patterns. DKGs will complement QKGs to explore poorly explored questions to new pattern discovery.
-
- **Animation Space** will connect open-source visualization software to the exploration of open-ended models to make the whole search transparent, highly visual and reproducible (Figures 1 and 2a, visualization layer).
- **Reporting** will develop a procedure to automatically explain the structure of the Bayesian space modeling module. It will also communicate the module using visualizations of the procedures followed by the Universal ETLs and Bayesian space models modules (Figures 1 and 2a, reporting layer).
- Robhoot 1.0 testnet is an automated reporting generation on "Biodiversity, Global Change and Sustainability Research" to explore the robustness of the automated research cycle accounting for **Universal ETLs**, **Bayesian space models**, **Animation space** algorithms and **Reporting** in natural processing languages.

- **Tools and Methods**: Multilayer networks metrics, Bayesian Networks, Julia, Python, Open-source software protocols, Gitchain, ETLs open-source software, Kafka, Clickhouse, Fluentd, Hadoop.

- **Novel territory**: Develop universal open-source ETLs algorithms and Bayesian space models and connect them to reporting automation in a "Biodiversity, Global Change and Sustainability Research" case study.

3.2 *ROBHOOT* v.2.0: Causal Knowledge Graphs

Rapid methodological advances is making AI more comprehensible (i.e., Explainable AI (refs)) but still not sufficient to fully decipher the processes underlying the empirical patterns. This limits the connection between prediction and knowledge power (Figure 2) (refs) critical to evaluate blackbox predictions match the plausible causal mechanisms predicting the data. *ROBHOOT* v.2.0 will be achieved by developing novel causal inference methods integrating neuromorphic computing in neural networks of spiking neurons (Noise as a resource for computation and learning in networks

Word	Meaning
Reproducible knowledge graph	High-resolution tracking of the research cycle to make it fully open and transparent
Evidence-based knowledge gap	Solid scientific knowledge facing constraints to be transfered to benefit society
Research-based knowledge gap	Differential access to the research knowledge limiting information transfer to the society
Discovery-based knowledge gap	Automated exploration of novel scientific paths to provide multidisciplinarity solutions to global sustainability problems
Question-based knowledge graph	Poorly explored questions important for multidisciplinarity solutions to global sustainability problems
Data-based knowledge graph	Poorly explored data integration important for multidisciplinarity solutions to global sustainability problems
Automation	Functionally interdependent algorithms targeting minimal human-driven interference
Knowledge-inspired society	Open access reproducible reports for global society
Neutral-knowledge generation	Open reproducible reports accounting for the many biases of the scientific process

Table 1: Glossary of terms.

of spiking neurons) into Bayesian inference (**WP5: Bayesian Spiking Networks (BSN)**. Inference is a complex problem especially in multidimensional landscapes (Inference in High-Dimensional Parameter Space; The frontier of simulation-based inference). In order to make inference from complex data more robust we will contrast predictions from Bayesian Spiking 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)**. The Bayesian space models module will ensure the search, the evaluation of models, trading-off complexity, fitting to the data and quantify resource usage (Figures 1 and 2a, inference and validation layers).

- Implementation of algorithms tracking paths of the research cycle with Reproducible Knowledge Graphs (KGs) (Figure 2b).
- Robustness and stability of searching and fitting procedures following a suite of open-source lineage client-tracker algorithms.
- **Tools and Methods**: Reproducible knowlegde graph algorithms and open-source packages (i.e., Renku and others)
- **Novel Territory**: Contrasting a set of Reproducible Knowledge Graphs algorithms to quantify the reproducibility, reusability, and recovery properties of the full research cycle.

3.3 *ROBHOOT* v.3.0: Federated Networks

- Deployment of deep learning algorithms to sample paths of the research cycle to produce populations of Knowledge Graphs (KGs) (Figures 2a-c).
- Exploration of the robustness of the automated research cycle combining optimization algorithms and the population of Knowledge Graphs (Figure 2c).

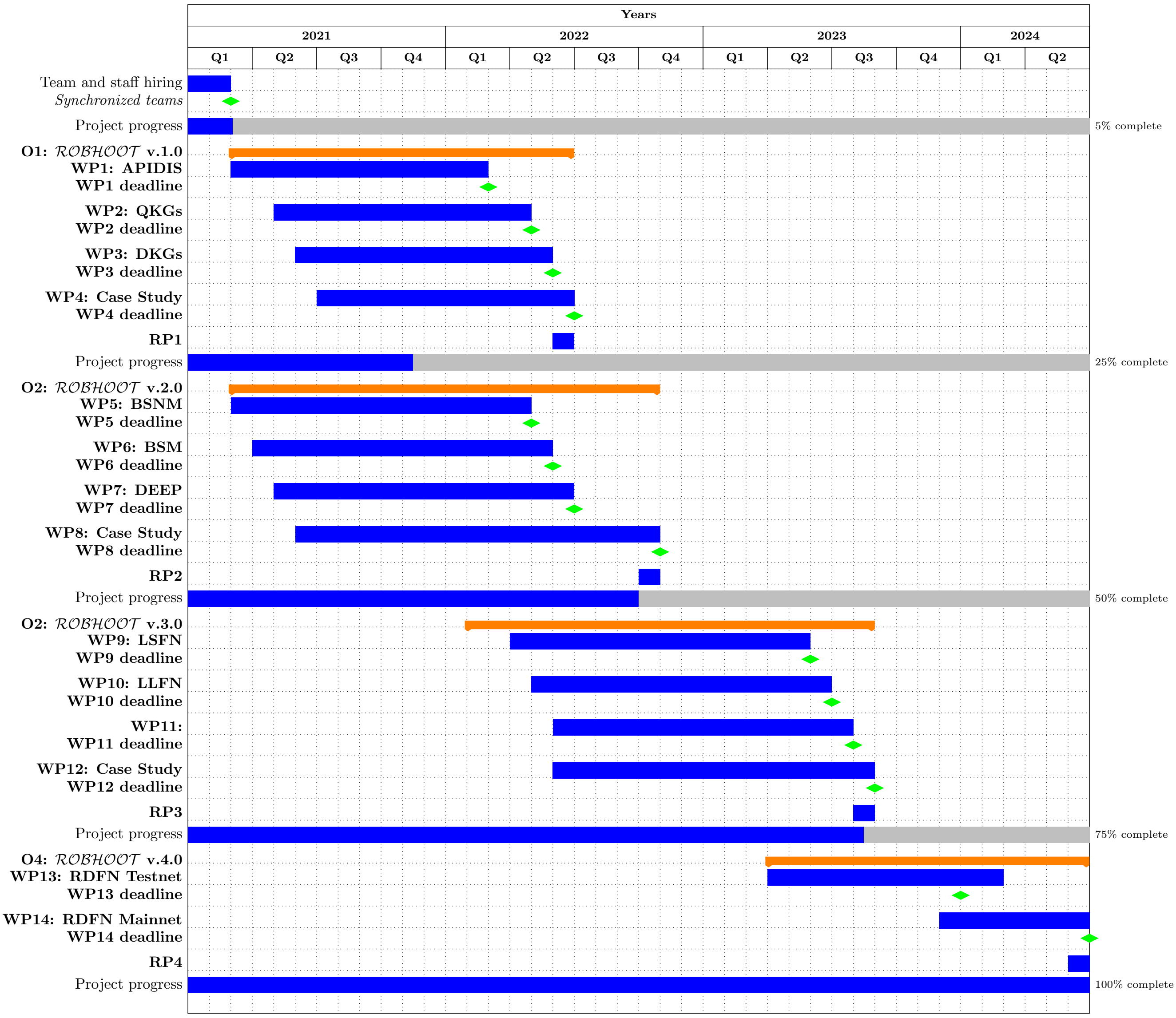

Figure 2: Roadmap: *ROBHOOT* **v.1.0** working packages **WP1** to **WP4** will take care of the integration of **Universal ETLs**, **Bayesian Space Models**, **Animation Space**, and **Reporting** to automate fully the research cycle (Figure 2a). *ROBHOOT* **v.2.0 WP5** to **WP8** will deploy knowledge graphs (KGs) into a fully traceable research cycle (Figure 2b). *ROBHOOT* **v.3.0 WP9** to **WP12** will explore deep learning networks to sample KGs populations to gain understanding of the robustness of the patterns in the data under distinct research paths (Figure 2c). *ROBHOOT* **v.4.0 WP13** to **WP16** will deploy KGs populations into a decentralized network of mutually trusting/untrusting peers with every peer maintaining the population of the KGs.

- **Tools and Methods**: Neural Biological Networks, Spiking networks, Bayesian Networks, Deep learning networks. Optimization algorithms.
- **Novel Territory**: Join Bayesian networks models to biology inspired deep-learning networks to efficiently explore constrained model space and the robustness properties of the populations of KGs along ensembles of the research cycle.

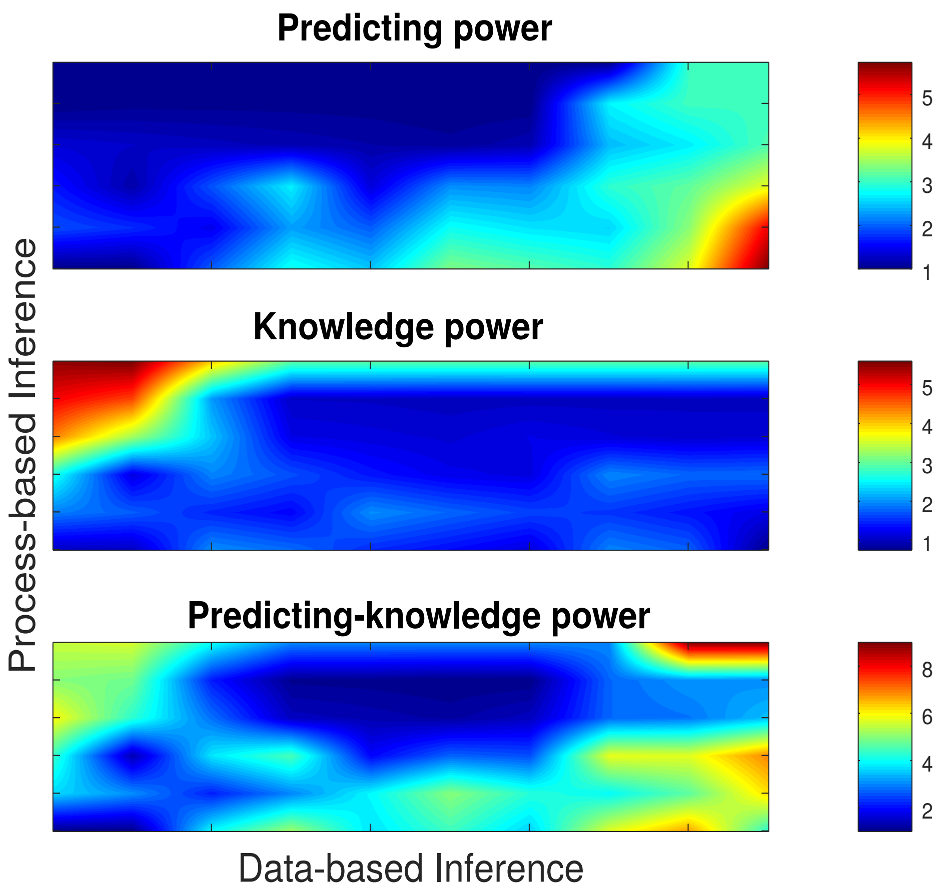

Figure 3: Prediction power (top), understanding (middle), and prediction-understanding power maps (bottom). x-axis represents data-based inference (i.e., gradient of AI methods from low (left) to high (right) predictive power). y-axis represents process-based inference (i.e., gradient of process-based methods from low (bottom left) to high (top left) understanding 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 understanding power map (middle) shows a red hot spot in the top left highlighting the region where the best mechanistic understanding occur. The predicting-understanding power map (bottom) shows the sum of the two previous maps highlighting a red hot spot where the best synthesis research joining predicting and understanding power of the empirical data might occur. The first research goal of this proposal aims to build an automated research platform to maximize the predicting and understanding power highlighted in the red hot spot of the predicting-understanding power map (bottom).

3.4 *ROBHOOT* v.4.0: Rapid Discovery in Federated Networks

- Deployment of a permissioned-permissionless distributed ledger technology to guarantee decentralization, open-access, neutral-knowledge-based network generation and prior confidenciality/posterior reproducibility of the KGs populations (Figures 2c and 2d).
- Exploration of a suite of consensus algorithms and smart contracts among trusted-untrusted peer-to-peer interactions to infer macroscopic metrics of the open research network (Figure

2d).

- Quantification of metrics to study the scalability-security-decentralization trade-offs when storing KGs in the research network (Figure 2d).
- Testnet case study to explore the interaction between consensus protocols and the scalability-security-decentralization trade-offs when committing the KGs to the distributed ledger.
- Mainnet to cryptographically link each population of KGs to previous KGs-ledger to create an historical KGs-ledger chain that goes back to the genesis ledger of the open research network. Launching of the mainnet to connect multiple database integration with real-time open-access citizen data science and knowledge-inspired societies.

- **Tools and Methods**: Distributed computing algorithms, Blockchain and consensus algorithms, BighainDB, Gitchain. Telegram open network, Golem.

- **Novel Territory**: Deployment of contrasting functional consensus algorithms to explore decentralization and robustness properties of the KGs populations along ensembles of the research cycle space.

Feature	*ROBHOOT*
Long-term vision	Global open-access to a fully reproducible knowledge-generation inspired technology
Breakthrough scientific and technological target	Collapsing evidence- and research-based knowledge gaps for a sustainable knowledge-inspired society
Novelty	Science-based technology emerging from targeted algorithmic discovery at the interface of multilayer networks, knowledge graphs, deep-learning, and consensus mechanisms
Foundational	Neutral-knowledge inspired technology for an emerging open science of science and science-society research disciplines
High-risk	Adapted to explore new terrirories into the open-science-technology-society interface ecosystem
Interdisciplinarity	Hybridizing expertise from distributed computing and deep learning to multilayer networks and the ecology and evolution of natural and digital ecosystems (Table 1)

Table 2: *ROBHOOT features along its developmental stages.*

The science ecosystem currently lack technologies fully automating the research cycle into the open-source digital ecosystem. Despite public institutions are demanding more reproducibility and openness of the data and the scientific process, and overall a shifting towards open and reproducible scientific and engineering landscapes, there are not currently open and integrated technologies aiming to compactly facilitate and distribute the scientific and engineering knowledge in open, reproducible and immutable knowledge networks (Tables 1 and 2).

Automating knowledge-generation requires the integration of many distinct features. Usually, knowledge-generation comes from interactions within- and between-layers of the scientific process (Figure 2a). The feedbacks occurring within and among layers in the science and technology ecosystem also provide unexpected behaviors that are difficult to anticipate. Therefore many feedbacks and interactions within- and between-layers are not easy to reproduce if not properly accounted for. We will take advantage of the open-source software community to explore knowledge graphs, optimization, automation, and decentralization algorithms together to study the robustness and reproducibility properties of the scientific process (Figures 1 and 2).

One way of visualizing the dimensionality of *ROBHOOT* in the digital ecosystem is to connect each layer of the scientific process (Figure 2a) to open-source software to gain functionality of the open research network (Figure 3). For example, Node 0 (left column, Figure 3) can be the Data Integration layer in Figure 2a. This node is connected to seven nodes representing open-source ETLs open-source software (i.e., central column, Figure 3). Connections between Node 0 and nodes 5, 6, 8, 9, 10, 12 and 13 can be rapidly evolving (i.e., indicated by the different red tones of the connections). Indeed, open-source ETLs are rapidly evolving towards accounting for many heterogeneous aspects of data integration (i.e., formats, historical-real time, storage, dimensions, size, bias and spatiotemporal resolution). ETLs can also be connected to a gradient of reporting generation (i.e., right column, Figure 3) noting reports containing only a subset of the interactions of the digital ecosystem network. The network of the fully automated research cycle can be one where Nodes 0, 1, 2, 3, and 4 represent the different layers of the research cycle (left column, Figure 3 and Figure 2a) connected to the open-source software of the digital ecosystem (central column, Figure 3) to generate full populations of reports (right column, Figure 3).

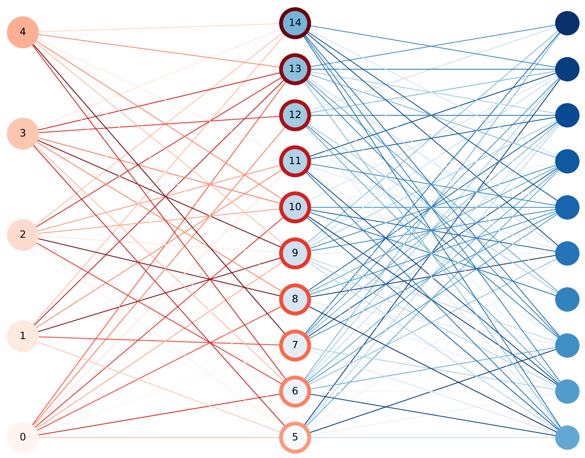

Figure 3: Robhoot in Digital Ecosystems: **Left column**: *ROBHOOT* **v.1.0** representing the research cycle as nodes from number 0 to 4: Data integration (0), Complexity Reduction (1), Inference (2), Validation (3), and Visualization(4)). **Central column**: Nodes representing the research cycle in the left column are connected to open-source software in the digital ecosystem. Connections with node number 0 in the left column can, for example, represent the ETLs open-source software interactions required to generate the **Universal ETLs** module. The same meaning applies to the different nodes of the left column. **Right column**: Each node represents a report meaning there is a reporting gradient generated by the connections to the open-source software from where each report is generated only using a subset of the research layers and open-source software.

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.

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)

Please provide, for each participant, the following (if available):

- 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
- a list 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)

Please complete, for each participant, the following table (or simply state "No third parties involved", if applicable):

- 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.).
- provide the documents that you need under national law(if you already have them), e.g.: an ethics committee opinion; the document notifying activities raising ethical issues or authorising such activities If these documents are not in English, you must also submit an English summary of them (containing, if available, the conclusions of the committee or authority concerned). 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

Please indicate if your project will involve:

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

References

[1] 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.

[2] Vlad Günther and Alexandru Chirita. " Scienceroot " Whitepaper. 2018. URL https://www.scienceroot.com/.

[3] 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.

[4] Tom E. Hardwicke, Maya B. Mathur, Kyle MacDonald, Gustav Nilsonne, 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.

[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] 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.

[7] Golem. The Golem Project Crowdfunding Whitepaper. *Golem.Network*, (November): 1–28, 2016. URL https://golem.network/crowdfunding/Golemwhitepaper.pdf.

[8] Nikolai Durov. Telegram Open Network. pages 1–132, 2017.

[9] 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.

[10] 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/.

[11] 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.

[12] J Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61: 85–117, 2015.

[13] 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.

[14] Yolanda Gil, Bart Selman, Marie Desjardins, Ken Forbus, Kathy Mckeown, Dan Weld, Tom Dietterich, Fei Fei Li, Liz Bradley, Daniel Lopresti, Nina Mishra, David Parkes, and Ann Schwartz Drobnis. A 20-Year Community Roadmap for Artificial Intelligence Research in the US Roadmap Co-chairs: Workshop Chairs: Steering Committee. Technical report, 2019. URL https://bit.ly/2ZNVBVb.

[15] Christian Steinruecken, Emma Smith, David Janz, James Lloyd, and Zoubin Ghahramani. The Automatic Statistician. pages 161–173.

[16] 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), 2020. doi: 10.1126/sciadv.aav6971. URL https://advances.sciencemag.org/content/6/5/eaav6971.