Iris.ai's project Aiur: an open, community-governed AI Engine for Knowledge Validation

JACOBO ELOSUA (CFO), ANITA SCHJØLL BREDE (CEO), MARIA RITOLA (CMO) and VICTOR BOTEV (CTO)
Co-founders of Iris.ai

Token sale white paper public version 1.2

The world needs science. Complex challenges ranging from climate change to preventive medicine require us to put our best minds together to solve them. And we do live in a world where more scientific knowledge is available to us than ever before - but the irony is that our politicians doubt its legitimacy, researchers often do not communicate beyond close academic quarters, university libraries cannot afford to pay looming subscription fees and publishing houses generate profit by keeping vital results hidden behind heavy walls, going after those who breach them with deadly force. In spite of Tim Berners-Lee creating the World Wide Web in order to share scientific knowledge, it seems we are only marginally closer today than we were back then.

Meanwhile, the research process is plagued with hard-to-justify inefficiencies, and among them, the growing need to distill and filter through all the noise. Much like Google became the solution to the static link lists, so we now need a machine-assisted system that takes us beyond a result list and into distilled, verified knowledge. This involves, among other things, challenging the old, ill-conceived power structure of publishing houses that hold knowledge hostage behind increasingly hard to justify paywalls. It also involves building a community-centered alternative to existing science governance approaches, leveraging the full potential of decentralization technology.

Some authors have argued that the reason nobody has adopted a distributed ledger at scale, ten years after it was invented, is because nobody wants it [Stinchcombe 2017]. We disagree. We side with those who believe that blockchain governance design is one of the most important problems out there today [Ehrsam 2017]. And one still requiring ample experimentation. Current governance models in the world of science do not work and, as smart contract technology evolves, ledgers, protocols and algorithms provide great possibilities to rethink current flawed practices. In this white paper we make the case for a truly decentralized, open approach to knowledge validation. We believe time is ripe for a functional, distributed, encrypted ledger to gain pivotal importance collectively organizing and fostering the advancement of existing and future scientific knowledge.

To this end, at Iris.ai we propose building Aiur - a Knowledge Validation Engine. Unlike other projects to create exchange-inspired architectures where promoters sit on top [Dinkins 2017], we regard ourselves as an agent of a rich, horizontal, budding ecosystem fighting to democratize science, against many threats. Here we present a technology platform and governance structure leveraging the blockchain for both AI-contributor and AI-user flows, transparently, accountably and in scale. The AIUR tokens introduced in this white paper will run on top of the Ethereum platform in two distinct project stages. Through an initial token sale we target raising c. EUR 10,000,000 to build a fully decentralized, community-governed Knowledge Validation Engine, in line with Iris.ai's mission and long term roadmap towards building the world's leading machine-driven science assistant.

Categories and Subject Descriptors: CS [**IS**]: DM—*DA*

General Terms: blockchain, token sale, artificial intelligence, scientific research, systematic mapping studies, Knowledge Validation Engine

■

1. CURRENT CHALLENGES FACED BY SCIENCE

Scientific researchers, in their different shapes and forms - working at R&D departments, research institutions, innovation labs, higher education institutions, and also students - currently lack adequate software tools to process effectively the vast, fast-growing body of scientific knowledge being researched and generated across disciplines worldwide.

To date these tools have mainly been offered by the long-established, dominant players in the scientific publishing industry and by the world's largest software companies. Neither set of tools has placed researchers' key interests at heart in their development. In fact, publishers focus their services on search services and the provision of statistics, without tailoring tools to the actual full research process. As a result researchers today struggle finding the right factual base to serve as the corner stone of their own research efforts. Uniquely poor historical competition dynamics in the professional publishing space [Forgues and Liarte 2013] have resulted in both fragmentation of access to knowledge, and a startling disincentivization of any third party led disruption efforts attempted in the past. Software giants, on the other hand, have approached this space, but they have different goals and priorities leveraging their giant scale. They have consistently treated research as a side market, developing general tools, again not taking researchers' full process into account.

Looking in more depth into the current woes faced by scientific researchers [Julia Belluz and Resnick 2016], in this paper we place the focus on five key fronts: (1) information overload; (2) access barriers; (3) reproducibility issues; (4) built-in biases; and, (5) incentive misalignment.

These unresolved issues deeply affect the quality of research published. Unless they are effectively addressed we risk missing crucial problem-solving opportunities as a human species, thus wasting an unprecedented amount of the most valuable type of capital available to us - science-backed knowledge.

1.1 Information overload

We face a pressing societal global problem: the amount of scientific information we have as a human species is unprecedented and growing. No human mind can cope with the vast volume of research being generated today. Some estimates quantify the current rate of research publication at over 4,000 new papers per day [Jinha 2010]. Furthermore, in this world of abundant knowledge as much as 50% of papers published in some fields are read by less than three people [Eveleth 2014], meaning that the cutting-edge research output being generated by some of our brightest minds in academia is currently not being effectively deployed. This unmanageable information overload slows down and introduces massive inefficiencies in both academic and corporate research processes, hampering global innovation. Amid the vast, fast-growing volume of published re-

search produced today, finding the gem articles that can solve our problems, has become impossibly challenging.

1.2 Access barriers

Traditional publisher business models are coming under increased scrutiny [Vogel 2017]. The sustained, abnormally high relationship between economic returns yielded and business risks assumed by these legacy models has faced harsh criticism from scientific researchers, academic institutions, policy-makers and the general public alike.

And for a good reason [Buranyi 2017]:

> ”Scientists create work under their own direction funded largely by governments and give it to publishers for free; the publisher pays scientific editors who judge whether the work is worth publishing and check its grammar, but the bulk of the editorial burden checking the scientific validity and evaluating the experiments, a process known as peer review is done by working scientists on a volunteer basis. The publishers then sell the product back to government-funded institutional and university libraries, to be read by scientists who, in a collective sense, created the product in the first place.”

Meanwhile, the global Open Access movement has made considerable strides is fulfilling Tim Berners-Lee's vision for The Next Web of Open Linked Data. By way of example, in 2015 the number of scientific papers available through CORE, arguably the world's largest Open Access research repository, stood at 25 million. This number has grown by c. 83% p.a. to reach over 85 million at year end 2017 [Editors 2018].

1.3 Poor reproducibility

Users, readers and developers of scientific knowledge have documented a surprising deficit in research reproducibility [Hutson 2018]. This is one often cited example [Collaboration 2015]:

> “Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. Replication effects were half the magnitude of original effects, representing a substantial decline. Ninety-seven percent of original studies had statistically significant results. Thirty-six percent of replications had statistically significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; 39% of effects were subjectively rated to have replicated the original result; and if no bias in original results is assumed, combining original and replication results left 68% with statistically significant effects. Correlational tests suggest that replication success was better predicted by the strength of original evidence than by characteristics of the original and replication teams.

Substandard reproducibility of published research studies adds to pain points suffered by students, researchers and R&D departments across sectors. And when considered in combination with other problems outlined in this section, reproducibility deficits make it fundamentally hard to build new knowledge on top of old results. Science is a complex system, with many components intertwined and linked together. If original evidence is not perceived as solid, the possibilities of new research approaches failing become considerably higher.

1.4 Built-in biases

Existing tools focused on scientific search have been built with a common keyword and citation-based architecture that incorporates serious issues with learning-over-time and the identification and address of biases. Keywords are subobtimal building blocks of knowledge, originating from a time of computation power scarcity. Citations, on the other hand, incorporate counterproductive biases in the overall body of science, often through the all-too-human dynamics of academic advancement. This common architecture deployed also limits the user to fields already known, hampering interdisciplinary discovery. Current search engines have been devised with the underlying assumption that a user knows what she is looking for. They do not cater for a use-case where a researcher is not aware of the precise terminology employed in a new field of inquiry. This is to say that today's products cater for known-unknowns, but fail at helping researchers deal with unknown-unknowns, thus artificially siloing information in narrow specialist domains. We need more democratic governance of the world of science, with more equal rights for all researchers to publish ground breaking discoveries, overcoming known existing biases (i.e. age, demographics, background, etc.).

This existing biases can be compounded by the new stack of machine-powered science discovery technology. As highlighted among others by Yoshua Bengio [Pearson 2016], we need to help make the inner workings of AI algorithms more transparent and accountable, fighting citation-based and other well-researched cognitive biases.

1.5 Incentive misalignment

Research professionals - 11 million according to our calculations - are currently forced to deliver, publish and review on tight deadlines, with little to no accountability and reward for authors and reviewers, creating perverse incentives towards exaggerating facts and omitting assumptions and constraints. In fact, experiments that produce null results have been shown to face a higher barrier to publication than those that yield statistically significant differences [Annie Franco and Simonovits 2014], evidencing a gross misalignment between the actions undertaken by different individual actors in the space and the body of science considered as a whole. Quality of research suffers as a direct result of these practices.

The challenges outlined above, when combined, mean that professionally reviewing an article has become really hard nowadays. A new generation of Artificial Intelligence-powered software tools could provide a reviewer with the most relevant information needed to understand the work at hand, whilst also helping with the pre-identification of any potential reproducibility issues. We view this as an optimal way to establish a modern, high quality review process that can scale and a fairer, more transparent and more unified across fields one too.

Furthermore, building the software tools researchers demand requires being able to tap into a sufficient volume of high quality training data at affordable costs. Despite some initiatives in this direction [Suleyman 2017], we suspect the bigger players in the field of AI will not, left alone, provide the wider community with this much needed abundance of suitable training data on competition-enabling terms. In fact, we aim to address some of the admonitions

put forward, among others, by Jon Evans [Evans 2017], reflecting on how:

> "The pendulum has already begun to swing back. Big businesses and executives, rather than startups and entrepreneurs, will own the next decade; today's graduates are much more likely to work for Mark Zuckerberg than follow in his footstep. Evans goes on to say that AI doesn't just require top-tier talent; that talent is all but useless without mountains of the right kind of data. And who has essentially all of the best data? That's right: the abovementioned Big Five [Alphabet, Amazon, Apple, Facebook, and Microsoft], plus their Chinese counterparts Tencent, Alibaba, and Baidu.

We see this development as problematic, posing a grave threat to the dynamism of startup-induced, bottom-up free market competition, with the resulting negative effects on innovation and growth. We believe as well that small players' ability to compete and contribute value to the future development of AI hinge on the democratization of access to scalable, unbiased, high quality annotated data, particularly in the short term.

This is not to say that access to data is the silver bullet solving, in one go, all issues affecting AI development competition. The algorithmic front and the hardware front exist too. But fair, agent-neutral access to needed data is a critical issue that needs to be addressed from a community-centered point of view.

2. THE BLOCKCHAIN AT PLAY

Objective truth should be decentralized. In fact, scientific knowledge is arguably the ultimate decentralized system, particularly as we have transitioned from analogue into digital. By design and intention, the body of information forming scientific knowledge has the following characteristics: (1) in essence, it is not controlled by a central agent and is, by and large, individual node-independent; (2) through exposure to public scrutiny and constant challenge, no single contributor can appropriate any part of its value durably; (3) it is highly reproducible at theoretically close to zero marginal costs; and, (4) it is preciously valuable for a large and fast growing cohort of current and future users. Unfortunately, its actual development praxis has been severely skewed, distorting some of this 'natural' characteristics over the years. We fully share the view that successful deployments of the blockchain should be measured against the standard of achieving something meaningful for society [Buterin 2017a] [Johnson 2018].

An open, scalable, decentralized platform for knowledge validation -a Knowledge Validation Engine- offers an optimal way to fix distortions that challenge the ground-level effectiveness of scientific knowledge generation and dissemination world-wide, returning trust to the system. A community-run engine capable of checking the underlying factual base of a given input text, provides us with a unique opportunity to unbias our entire knowledge base, and doing so through a new prism built with the highest transparency and accountability standards.

And building this is today possible because Bitcoin's revelation has been profound:

> "It has shown that it is possible to use a network of computers, connected via the Internet, to build and maintain a set of valuable shared datain this case a ledger of account balances that prevents counterfeitingwithout the need for a trusted authority. Think about that: from a bunch of anonymous computers that have no reason to trust one another, an iron clad network has emerged that can support a whole currencyliteral money, what could be a more valuable target for hacking or compromise? And yet there it stands, unperturbed amid the chaos of the Internet.

As pointed out by Adam Ludwin [Ludwin 2017], a decentralized application allows you to do something you can already do today but without a trusted central party:

> "There remain question marks over whether decentralized applications will actually be useful to most users relative to traditional software. In fact, on almost every dimension, decentralized services are worse than their centralized counterparts:
> —They are slower.
> —They are more expensive.
> —They are less scalable.
> —They have worse user experiences.
> —They have volatile and uncertain governance."

And no, this isn't just because they are new. This won't fundamentally change with bigger blocks, lightning networks, sharding, forks, self-amending ledgers, or any other technical solutions. That is because there are structural trade-offs that result directly from the primary design goal of these services, beneath which all other goals must be subordinated in order for them to be relevant: decentralization.

Whilst disagreeing on the immutable nature of any scalability, user experience and governance shortcomings faced by blockchain developments, we fully agree with Ludwin on the speed and cost ones. And more importantly, we share the viewpoint that censorship resistance, a very present issue in the current world of research, is a pivotal feature underpinning justified blockchain deployments. In fact, as a researcher, if you are not able to publish your work to a well-known conference or journal, a process currently plagued with systematic biases, your research risks not being found by search engines, constituting a modern age form of censorship.

But beyond censorship resistance, at Iris.ai we believe that knowledge validation offers fertile ground for the development of open, transparent and accountable software, adding significant value to multiple stakeholders and users along the way. Intertwining blockchain mechanics, for example decentralising and opening up how machine algorithms are fed factual data, presents great impact potential for humanity as a whole. We believe it should result in better quality algorithms with greater traceability, removing of conscious and unconscious biases in how we build the datasets used to teach machines how to understand fact-based reasoning.

Whilst many blockchain developments have focused on anonymity as a key underlying feature, we are attracted by the flip side of full anonymity. A flip side that, ironically, only distributed ledgers can empower today: full scrutiny. Entity-independent trust fuels this new brand of scrutiny that, in our view, should power how scientific knowledge is organized and advanced forward in the current digital era.

The emergence of this full scrutiny paradigm demands twin developments: technology developments, on one hand, and governance developments, on the other. Uniquely, both required developments can be attained through the design and implementation of smart contracts. In this white paper we present our initial interpretation of how to leverage smart contract technology to build an open, community-governed AI Engine for Knowledge Validation.

As a novel experiment in the blockchain space, we will place a bet in the development of community-oriented taxation systems -

where each transaction will be treated individually as a result of the application of transparent criteria - to reward and penalize different behaviors towards the Aiur ecosystem.

3. PROPOSED SOLUTION

The proposed Knowledge Validation Engine should go a long way to help scientific researchers address the current key issues they face outlined above (see 'Current challenges faced by science'):

—Information overload. Directly, signalling which papers' factual content structure can be validated (or not), hence enabling powerful new filters. And indirectly, empowering Iris.ai and other product developers to build better artificial intelligence science discovery software.

—Access barriers. Directly, Through a new open access research repository contemplated in the development roadmap, without charges to users for reading content (but restrictions to host or redistribute). And indirectly, pressuring traditional publishers to subject their content to additional scrutiny.

—Reproducibility issues. This issue should be addressed directly, identifying issues in papers that might compromise the ability to reproduce the results of the experiments carried out and reported by a paper's authors. And indirectly too, raising the bar of expected public scrutiny.

—Built-in biases. We see this as a long-term process, with Aiur users gradually realizing that there is a large volume of highly relevant results de facto not visible through existing search engines. Additionally, Aiur should provide an invaluable control set to test the effectiveness of the citation system and help newly written articles have a more unbiased, comprehensive and easy to build citation list.

—Incentive misalignment. As cornerstones of the envisaged nascent community, authors will obtain token rewards for embracing increased openness standards, including comparable rewards for publishing failed results. Aiur will also open up the peer review process, and it will facilitate the generation of specialized datasets to develop and/or test models with a supervised layer.

As mentioned above (see 'The blockchain at play'), building Aiur and the targeted ecosystem around it requires twin technology and governance developments. On the technology side, smart contracts will establish an Institution regulating: (1) how tokens will be generated, making valuable contributions to the platform; and, (2) how tokens will be used, tapping into the platform's algorithmic brain. This Institution will set the system's policies relying on an Oracle. The Oracle will make external market readings to set a rate between AIUR and ETH and compute the minimum viable transaction limit and a few additional properties, like applicable taxation level (see 'Key policy mechanics' section below).

On the governance side, Iris.ai and Aiur will be two different entities. We will define smart contracts to enact a Constitution, regulating: (1) how the ecosystem will function initially ('Phase 1'), until the earlier of the initial stability targets being reached with the community supporting the transition or the 18 month token generation event anniversary backstop; and, (2) community member rights and obligations, consensus building and decision making mechanism, particularly relevant post transfer of full control to the community, when DAO governance standards will be proposed for adoption and Iris.ai's tokens will be redistributed and/or burnt to bring its stake down to the maximum cap set ('Phase 2').

We believe the kind of centralized trust model put in place during Phase 1, where Iris.ai will act, in essence, as project lead and core developer (i.e. in a role akin to that of a service contractor), will be not only useful but absolutely required in the project's early stages, whilst acknowledging, at the same time, that it clearly would not be sustainable in the long term.

In a first instance the proposed decentralized Knowledge Validation Engine - Aiur - will be separately built on top of Iris.ai's existing products, and in particular the company's current AI training platform. In this platform users are currently able to select a content item (i.e. a scientific paper URL), and proceed to highlighting the words or short phrases that best capture the essence of the piece of content, thus helping improve natural language processing algorithms. Each suitable training input provided will generate tokens for the respective submitting AI contributor; and, at the same time, tokens will be deployed to tap into the algorithmic services provided by the engine to AI users (see section Supply and demand policies' below for further details on the token mechanics).

The full development roadmap envisaged is discussed below (see section 'Aiur development roadmap'). Once built, this proposed Knowledge Validation Engine would address the issues faced by science listed above (see Current challenges faced by science). Its use-cases include, for example, having a decentralized system tell any aspiring user if the presented research is reproducible and how to go about it; enabling a more effective deployment of science by non-researcher individuals and smaller organizations [Vishnefske 2016]; inspiring community members in the development of new tools and services on top of better validated research; or, laying a solid foundation for the future development of AI with a more conscious treatment of bias - one that can incorporate bias mitigation development loops into its core. Over time we anticipate that the proposed Knowledge Validation Engine's core functionality will be a critical component in developing, among other solutions, Iris.ai's fully fledged AI Science Assistant.

We choose to build the Aiur Knowledge Validation Engine on top of Ethereum because this network is open-ended by design; supports smart contracts; enables the interaction with market forces through publicly traded tokens; and, through its community's efforts, we believe it will continue to be extremely well-suited to serve as a foundational layer for a very large number of both financial and non-financial protocols in the years to come [Jason Teutsch and Brown 2017a]. Ethereum is not designed for a specific application but rather as a platform to build applications that can execute arbitrary code, i.e. smart contracts. A smart contract uses software code to implement human intentions by dynamically carrying out instructions embedded in tokens associated with a contract, rather than relying on legal texts interpreted by courts, regulatory bodies or other legal institutions. In similar fashion to how Bitcoin was ideated to address vulnerabilities in the current banking system, we believe the proposed Aiur Knowledge Validation Engine can play a big role in forming an ecosystem around it to open up and democratize effective access to science for everyone.

We believe our technology stack and product versions already released place us in a unique position to kick-start and initially lead the creation of a public-facing, shared and scalable factual data power engine. We aim to improve researchers' processes through the collective development of text understanding machine intelligence. And to do so placing the generation of broadly distributed societal value at its very core.

In our view, the creation of the Aiur Knowledge Validation Engine should contribute durable value to be captured by a broad community of crypto token holders. With first mover advantage and community-ownership dynamics in its DNA, the project will

provide an open platform to generate critical inputs in the development of high-value-added, continuously improving AI-based products and services.

Universities, research institutes and RD departments spend USD 128 billion a year on digital enablers, and a medium sized department can save millions yearly using new generation knowledge discovery tools. The Aiur Knowledge Validation Engine should both drive significant efficiency gains in current work flows, as well as help spot attractive new research-linked revenue opportunities. These organizations have their own internal tools and processes, and would connect these directly to the Aiur API. They would pay AIUR to query the engine.

A variety of future applications would rely on the Aiur Knowledge Validation Engine. This would tap into markets such as patent writing and prior art searches, hedge fund technology predictions, research funding and venture capital in addition to a range of not-for-profit applications such as tracking corporate research for marketing. Third party tools could charge their clients fiat or crypto, and then pay AIUR for the querying of the engine.

All payments for services from the Knowledge Validation Engine will be done to the Aiur Financial Institution. At the same time, new token issuance to community members will be strictly restricted to verifiable value contributions. The Aiur Financial Institution will manage AIUR demand and supply flows, burning excess tokens accumulated via a sustained influx of capital. These mechanics will govern the value growth of the community.

In a very long term perspective, teaching an AI to understand the world around it through scientific knowledge we believe will be an important aspect to building artificial general intelligence - another reason why true community governance is so vital to Project Aiur.

4. PLANNED ECOSYSTEM

In our opinion new ecosystems need to be built for the collective exploration of alternative AI development architectures. The proposed Knowledge Validation Engine, Aiur, provides an optimal opportunity to structure such an ecosystem, with a set of different agents, including researchers and engineers, committed to developing technology and products openly as part of a community with a common purpose.

In this ecosystem, forged around a shared vision - to democratize access to and extend the reach of scientific knowledge world-wide -, we envision at least four profiles contributing value to the design and development of the proposed Knowledge Validation Engine: (1) AI trainers; (2) coders; (3) quality assurance; (4) researchers and reviewers. At the same time, we contemplate four basic user profiles leveraging Aiur: (1) software developers, both commercial and open source; (2) R&D departments and research institutes; (3) academic research departments and consortia; and, (4) individual researchers.

Without turning a blind eye to potential risks associated to any future misuse of sufficiently advanced machine intelligence, we firmly believe that humanity at large will be able to extract value from the technology developed. And we state this considering both, any knowledge-thirsty human individually considered, and society as a whole benefiting from the positive economic externalities associated to a more developed, open knowledge bank.

Iris.ai's initial contribution to the creation of this desired ecosystem will come through the adaptation of our current AI training platform, and including new token generation and token use mechanics (see 'Aiur development roadmap'). In this AI training platform, currently used by c. 9,000 individual trainers world-wide, any user can register and train datasets to provide a baseline against which to improve our natural language processing algorithms. Logging these trainer inputs in a blockchain and incentivising a large volume of potential trainers to annotate additional datasets, we believe will contribute to the creation of invaluable, high quality datasets for the future development of better text understanding machines.

Beyond our own contributions as the platform's original promoters and a significant future user (see section Supply and demand policies below for further details on the expected role of Iris.ai as a platform user), our vision includes a wide array of market and non-market participants building on top of and leveraging a publicly available Knowledge Validation Engine tool to advance machines' understanding of scientific knowledge from a diversity of complementary angles.

As mentioned above (see 'The blockchain at play' section), we envision two distinct project phases, with Iris.ai playing critically different roles in each of them. During Phase 1, Iris.ai will use funds raised through the initial token sale to kick-start the platform development process. This phase will include taking the lead in smart contract creation, software design and development, communication and community building, and general project management activities. In a maximum 18 month timeline following the initial token generation event, Iris.ai will make every effort to launch and stabilize a precursor Knowledge Validation Engine platform. Once this goal is achieved, as established through a combination of objective criteria met and a confirmatory community vote, Phase 2 will kick in. In this second phase Iris.ai will relinquish control over the platform to the initial community formed around the project. As per the Constitution provisions reflected in one of the seminal smart contracts put in place, Iris.ai will be on an equal footing with all other community members in defining and deciding the future evolution of the platform as a truly open and decentralized one.

At transition from Phase 1 to Phase 2, the amount of tokens held by Iris.ai will be reduced from 50% plus one to the general maximum cap. This reduction will be effected through burning and/or redistributing tokens to community members. Foundational community member status, acquired through participation in the initial token sale, will weigh in significantly at implementation. The universal ownership cap, set initially at 2% and subject to community revisions via voting, will be continuously monitored for enforcement via a smart contract.

Taking the cue from Vitalik Buterin [Buterin 2017b], and staying clear from solutions such as creating a self-destructing entity to build the proposed Knowledge Validation Engine [Shochat 2017], we postulate a centralized trust model as useful in the project's early stages (Phase 1), moving to well-defined multifactorial consensus once initial stability has been reached (Phase 2).

A vital part of every ecosystem are the rights and obligation that correspond to the community membership. We try to keep those simple, but still a minimum set is required and we outline that in the lines below.

Similarly to open source projects, Aiur is designed not be owned by a single entity or organization, but by a decentralized, distributed community.

This design gives each member of the community a number of fundamental rights and obligations. Community member status requires holding a minimum one AIUR token and signing a membership agreement.

4.1 Rights

Each community member has the right to:

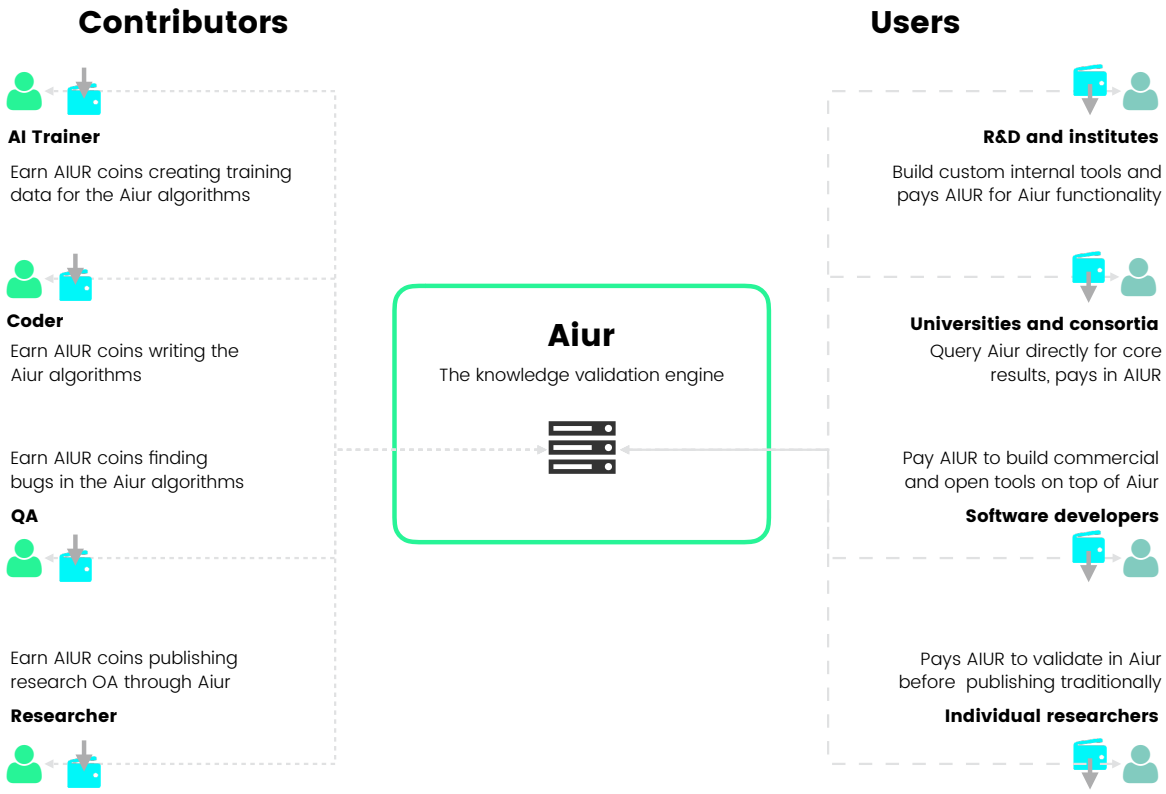

Fig. 1. The Aiur ecosystem

—Use the service of Aiur at any time without the need to host the system (R1).

—Download and host Aiur on their own premises (R2).

—Download, see, comment and modify the code of Aiur (R3).

—Submit code changes in pull requests for approval to the community (R4).

—Report issues and express opinion about the design, implementation and execution of the system (R5).

—Contribute in decision making for accepting issues and pull requests (R6).

—Propose publication of authored or third party open access scientific research, in whatever form that may take in the future [Somers 2018] (R7).

—Submit annotations to the system for their potential selection (R8).

—Transparently scrutinize annotated datasets (R9).

—Challenge the system's individual components, such as research papers or annotations (R10).

4.2 Obligations

Each member of the community should respect the efforts made by other community members for building, maintaining and supporting the Aiur platform. Each community member is obligated to:

—Pay the usage contribution fee set for the use of Aiur services. This contribution fee could be reduced if the person is actively hosting an instance of Aiur on their own premises (O1).

—Allow bandwidth of usage to anyone in the community whenever hosting the service on their own premises, with rate discounts based on serviced requests (O2).

—Disallow direct access to non-members of the community, enabling access exclusively through service layers that guarantee payment by the non-members of the contribution fee set (O3).

—Make authored research available open access, also publishing failed results with scientific value (O4).

—Always behave and specifically vote with the community's best interest in mind (O5).

—Respect and uphold the Aiur constitution in the face of identified threats (O6).

The enforcement initially will be done on a best efforts basis.

5. AIUR DEVELOPMENT ROADMAP

This section outlines the main research and development sub-projects and associated tasks that need to be undertaken in order to create the Aiur Knowledge Validation Engine. The roadmap includes milestones and a draft target timeline, based on a four-montly release cycle, that will be refined in conjunction with the newly formed community.

5.1 Overview

Aiur is a fully fledged technical platform that shapes, organizes, and augments the community. The platform contains governmental, financial and research community tools. At the core of the platform is a research process facilitation framework that has the goal of pinpointing what are the building blocks of a scientific text, what does the reader need to know in order to be able to understand the text, what are the factual sources on which the text is built, and what is the reproducibility level of each building block and of the overall scientific text itself. The framework takes a scientific document in the form of a scientific paper or technical report as an input and provides an analytical report of the knowledge that builds the document, its reproducibility level and the input's hypothesis tree. The hypothesis tree, where each child hypothesis is a prerequisite for the parent hypothesis, leads to the presented root hypothesis in the input document. The report will also show the support levels for each of the hypotheses in the hypothesis tree. All of this will be based on a knowledge database of scientific documents accessible to the system at any given point in time. In order to achieve this, Aiur will comprise of a number of subsystems. These include a knowledge validation system, an AI training platform, an infrastructure service, a dispute resolution engine, a distributed science repository and a smart contracts framework. All these sub-systems, along with their planned functionality and milestones, are covered in more detail below. The architecture of the full Aiur platform is displayed graphically in Figure 2.

5.2 Subsystems - setup and research tools

5.2.1 *Knowledge validation system (project Blackstone).*

Blackstone will be at the core of project Aiur. Its fundamental goals are for it to be a system that can make sense of a one scientific document and enrich its meaning, prepare the building tree of the document and validate its roots. To be able to achieve that project Blackstone will be capable of the following:

—Mimic human understanding of a scientific text and human hypothesis extraction, using available meta-data information, such as document fingerprint, keywords, and other contextual data.
—Identify the most relevant documents to a given starting document based on the content of the documents and a given document similarity metric.
—Link information from all related and relevant documents to a scientific document and identify only those fragments that contribute to the initial document context, using available topic information or other document clustering information.
—Summarize and index the information needed for understanding the input document.

Given the complexity of this sub-sytem we will break the details of its functionality into four distinct parts - Hypothesis extraction engine, Knowledge tree builder, Reproducibility engine and Validity engine.

Hypothesis extraction engine

Understanding the core essence of a research document is important for the future development of a Knowledge Validation Engine. It gives possibilities for creating causal connections between scientific articles and also finding similarities on a lower structural level (like addressing solution, method or argumentation levels). The aim is to do research and develop software that can extract a pseudo hypothesis/argument (e.g. problem - solution - evaluation - results) from text. This component will be used for tools such as summarization of information, suggesting similar articles based on the acquired hypothesis knowledge, identification of building blocks of an article, finding true references, etc.

Input: document text, document keywords, document topic information, document vector, word vectors

Output: hypothesis causality graph, modelling identified problems with proposed solutions, their evaluation and possible results.

Roadmap:

—MVP: Identifying the main problem and sub-problems discussed in the input document.
—Milestone 1: Identify the main building information for the argument in the input document. Main problem, solution, evaluation mechanism and results and their corresponding counterparts.
—Milestone 2: Using the main building information to build the hypothesis causality graph.
—Milestone 3: Include information from images, tables, graphs, etc. black data into the hypothesis causality graph.

While the text of any research paper is essential, 'black data' also holds core value to the user. Extracting and identifying these components and presenting to the user the ones most relevant to the hypothesis in the applicable location of the synthesized hypothesis will increase the user's ability to more rapidly consume large amounts of knowledge. This could be a very useful base for building visual summaries of the articles.

Knowledge tree builder

The goal of this module is to build the knowledge tree of a given scientific article.

Input: document text, document keywords, document topic information, document vector, word vectors, document repository.

Output: knowledge tree of the document with predefined depth.

Roadmap:

—MVP: Splitting the document into building blocks based on the meta-information. Assign to each block a related best document enriching the understanding of the building block. Utilization of the manual citation system.
—Milestone 1: Create a tree that starts from the initial document as a root and using its building blocks assigns other related building blocks from other documents until a stop condition is met. Such stop condition could be a cyclic loop, or that there are no good enough documents to relate to, etc.
—Milestone 2: Integrate the hypothesis causality graphs into building blocks. Separate and use that information during tree build-

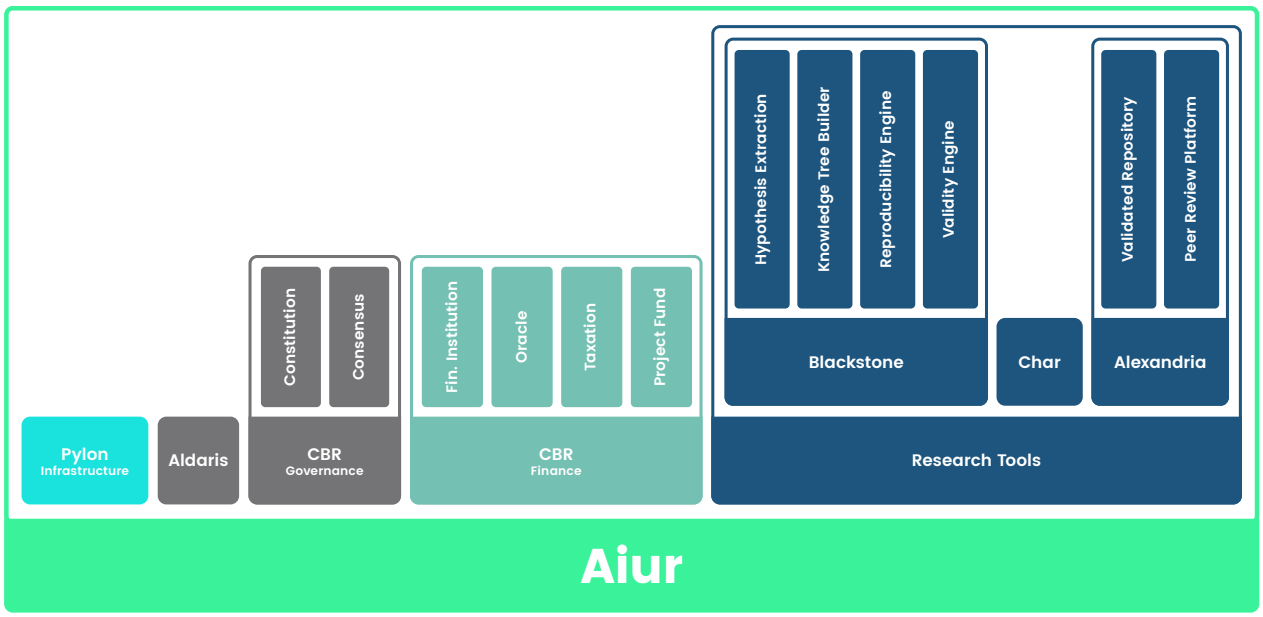

Fig. 2. The Aiur platform architecture

ing to make the knowledge tree as relevant as possible to the initial document.

Reproducibility engine

Input: document text, document keywords, document topic information, document vector, word vectors, document repository, hypothesis causality graph, 'black data'.

Output: reproducibility flag Red, Yellow, Green, and causes for the light signal.

Roadmap:

—MVP: Pure knowledge discovery of constraints and assumptions using them as causes for reproducibility issues.

—Milestone 1: Using 'black data' and definition of experiments for experiment reproducibility checks

—Milestone 2: For some areas, connect the engine to a simulation environment and test the reproducibility of experiments directly.

Validity engine

Input: document text, document keywords, document topic information, document vector, word vectors, document repository, hypothesis causality graph, 'black data', knowledge tree of the document.

Output: validity flag Red, Yellow, Green, with a full report containing the causes for the light signal based on the building blocks.

Roadmap:

—MVP: Checking the validity of the facts within the document itself for reproducible documents.

—Milestone 1: Analysis of building blocks and flagging for each branch of the knowledge tree.

—Milestone 2: For some areas, connect the engine to a simulation environment and test the validity of the branches blocks directly.

5.2.2 *AI training platform (project Char).*

A digital tool enabling both expert and non-expert trainers to annotate texts. The training platform has the purpose of helping the success of project Aiur. Main annotation requests will come from project Blackstone, but other projects might require annotation assistance as well.

Project Char will work on an on-demand basis the demand requests will be submitted by the execution team working on the corresponding requesting project, each demand request will be split into blocks, where each block will contain certain quota and when the quota is finished and/or a time gate is reached, new block should be opened before the annotation can continue. Blocks will be opened based on the requirements and schedule pre-defined in the demand request.

Since the annotations from Char will be used for AIUR token generation, certain quality control measures need to be ensured. The platform will log all annotation contributions transparently in the blockchain. It will also include algorithm for validating human annotations and unbiasing datasets.

To secure the annotation service from abuse additional security measures will be put in place:

—There will be time delay before consecutive requests, with minimum annotation time based on the text length several minutes).

—Annotations that are below certain algorithmic score will result in expelling the user from participation in the current annotation block.

—The annotations that fill in an active slot in the current block will not be announced until the full block is complete.

Roadmap:

—MVP: Tool for annotating document keywords and an algorithm for assessing keyword annotations.

—Milestone 1: Tool for annotation of document hypothesis and an algorithm to assess hypothesis annotations.

—Milestone 2: Expanding the MVP with functionality to balance the gathered dataset, identifying possible biases and adding an importance score to each new annotation.

5.2.3 *Distributed science repository (project Alexandria).*

The real value of validating research (project Blackstone - Sub-section 5.2.1) lies in actually using this research. Project Alexandria has the goal of acting as Aiur's library or archive of validated knowledge, where researchers should get royalties from their filtered participation and reviewers should get royalties for verifying research and keeping the high quality standards of the repository. Reviewers will get the necessary help from project Blackstone's results, and they will have the ability to raise issues for dispute in the Tribunal Service (Sub-section 5.3.1).

Besides the problems of validating published research and its quality, another problem that project Alexandria will address is the fact that all articles published nowadays in the main journals present algorithms and ideas that are always considered better than the current state-of-the-art. Project Alexandria wants to introduce as part of its distributed repository a Journal of Alternative Approaches (JAA) - those approaches still have to be innovative, show potential, and be backed up by facts and experiments as a regular scientific article, but do not necessarily need to be better than the state-of-the-art. Sometimes there are months of work involved in trying a certain approach that fails at beating the state-of-the-art. But that work should not be forgotten, left unpublished, or published only for a very narrow niche case. There is still value in publishing such results, because then other people will know such approach has been tried out before and how it has been evaluated and the reasons why it is not considered state-of-the-art. Having that in mind, system users can choose to go in that direction if they have a new idea on how to improve it, or not follow the given approach at all. That will also give more information about the approaches considered state-of-the-art, and reinforce why they are considered such and in what aspects are they better than other approaches.

Semi-automated review platform

Project Alexandria needs to provide a platform with tools facilitating the review process for articles to be included in the scientific repository. These tools should give the reviewer capabilities to:

—Access all relevant research needed to understand the text under review.

—Receive reports about the factual information within the text and also the knowledge it is based on.

—Access similar and contradictory hypotheses to the ones included in the text under review.

All community members can opt to be reviewers. The review process will have two branches - one through the semi-automated review platform, where a randomly assigned article is assigned for review based on a reviewer's expertise, and another through directly challenging the author via project Aldaris (Sub-section 5.3.1). Reviewers will receive tokens only if their review is not actively and successfully challenged during a predefined time frame by the author or a different community member.

Validated community-open research repository

The community repository should be distributed and hosted by everyone in the community. It should also contain all approved articles, sorted into several journals.

Authors will receive tokens only when their articles get accepted to the repository. The acceptance criteria will include no red flags generated from the service in project Blackstone (Sub-section 5.2.1) and at least ten successful reviews (an all green lights report from project Blackstone will count as one successful review). Additional token generation mechanisms could be devised by the community to incentivise derived impact.

Roadmap for project Alexandria:

—MVP: Journal of Alternative Approaches. First version of the semi-automated review platform. Consensus mechanism for awarding tokens to authors and reviewers.

—Milestone 1: Integrating tools from project Blackstone to improve the review platform.

—Milestone 2: Adding more journals with validated content.

5.2.4 *Infrastructural service (project Pylon).*

Project Pylon concerns itself with the necessary ground level infrastructure to make Aiur a reality.

Project Pylon will provide capabilities for open-community coding, testing and verification of the produced software, as well as orchestration of the necessary infrastructure for self-deployment, load balancing of requests and other services required for a fully fledged execution of the Aiur services. Project Pylon should be responsible for integrating a code repository of the other Aiur projects with the corresponding smart contracts and community mechanisms for governing the code base of Aiur.

Orchestration, self-deployment, load balancing and other smaller infrastructural service

Given the decentralized essence of the project, we want every user to be able to use the provided services. As a result, even if a user does not have the infrastructural capacity to execute and host the system at their own premise, they shall still be able to use it. This inherently means that project Pylon will provide orchestration containing the necessary scripts for setting up an infrastructure that can run the Aiur services on a local machine by the users or in a cloud provider (i.e. AWS, Azure, Google Cloud, etc.). Ideally for cloud providers we will use an infrastructure provider that accepts ETH/BTC or other cryptocurrencies as consideration for service payments, and which will allow for direct payments from the Aiur Institution. Self-deployment software will be incorporated in an Aiur master script able to download all Aiur projects, install their requirements, execute their tests and start the projects' services on

the underlying infrastructure.

Roadmap:

—MVP: Code base preparation, governing mechanisms for accepting new code and prioritizing new functionalities and bug requests. Building a reward mechanism for coders and bug finders.
—Milestone 1: Orchestration scripts for AWS. Self-deployment scripts for projects Blackstone and Char. Load-balancing service for all deployed instances. Discount mechanism for hosts of the system.
—Milestone 2: Expanding the functionality for other projects or services.

5.3 Subsystems - governance and financial tools

5.3.1 *Dispute resolution engine (project Aldaris).*

Project Aldaris has the goal of providing an effective dispute resolution mechanism between members of the community. It is specifically tailored to the issues related but not limited to:

—Project Alexandria (Sub-section 5.2.3) - disputes between reviewers and authors.
—Disputes between authors and the automated tools in project Blackstone (Sub-section 5.2.1).
—Disputes between coders and users or bug finders.
—Token generation event validity disputes in relation to project Char (Sub-section 5.2.2) and project Alexandria.

In general, the project will address operational disputes not directly related to the overall government and organization of the community, its members' rights and obligations, or the essence of the smart contracts governing Aiur.

Project Aldaris should operate as follows:

—Receiving a dispute inquiry. This communication should contain the reason for dispute and the bounty that is at stake.
—Notifying the challenged side.
—If the challenged side agrees with the argument put forth, the flagged issues should be corrected and no token transfers will be made.
—If the challenged side disagrees, then both sides stake their bounty to the Institution and, upon resolution, the winning party receives the bounty of the other party.
—To resolve the matter, Aldaris will invoke a community consensus mechanism to decide the winning party.
—Since running the dispute resolution process incurs costs, bounty money will be taxed.

Roadmap:

—MVP: Preparing an appeal service that allows for the resolution of disputes between reviewers and authors. A dispute will initially be in direct relation with a token generation event for a reviewer and an author. Any user should be able to dispute a review made by a reviewer and, if the dispute is successful, the generated tokens should be transferred to the community member questioning the review.
—Milestone 1: Expanding the MVP with functionality for coders and bug finders.
—Milestone 2: Broadening the general dispute resolution mechanism for any other matter potentially under dispute between community members.

5.3.2 *Smart contracts (project CBR).* Project CBR will consist of a set of smart contracts taking care of the initial crowd sale as well as all of the needed infrastructure ensuring a functioning and sustainable economy and community. It will facilitate payment operations, earning and spending of tokens, and overall utilization of the blockchain for achieving project Aiur's goals. Project CBR has two main branches CBR-Governance and CBR-Finance. CBR-Governance will deal with all governance related services - decision making, reaching consensus, constitutional rights enforcement and violation penalties, while CBR-Finance will take care of the community economy and provide services to maintain its future sustainability.

CBR-Governance

On the governance side, we will define smart contracts to regulate: (1) how the ecosystem will function initially ('Phase 1' - centralized trust), (2) how the transition will happen post 'Phase 1', (3) community member rights and obligations, consensus building and decision making mechanism, particularly relevant post transfer of full control to the community ('Phase 2' - decentralized self-organized community).

Constitution

Constitutional contracts will be responsible for:

—The transition between Phase 1 and Phase 2. This transition will follow the scheme described in Section 3, i.e. ensuring that either the initial stability targets have been reached or at the 18 month anniversary of the token generation event backstop. If one of those triggers is reached the Constitutional contracts will transfer all tokens beyond the 2% holding cap from Iris.ai's account to the accounts of other members, thus making Iris.ai an equal member of the community.
—Enforcement of community rights and obligations. When it comes to reversibility of actions, we believe in the merits of incentivising behaviours such as finding loopholes and their one-off exploitation, whilst, at the same time, we aim to penalize continued exploitation of system vulnerabilities made in bad faith. The initial mechanisms envisioned for penalizing undesired behavior by the community will be freezing of funds. In the future, penalization decisions and their framework will rest with the community .

Consensus and decision making

We have discarded *one token = one vote* and *one person = one vote* as either undesired or impracticable decision making criteria.

Aiur, particularly for the purposes of Phase 2 decision making, will rely on proof of value-added to award voting rights. Actions such as token generation and token use will be regarded as value adding.

Duration of time actively participating in the community, whilst in compliance with its regulations, will be treated favourably. For community members turned inactive, however, voting rights will decay over time.

Non verified identities will be penalized when it comes to collective decision making. Staking tokens will be required for participation.

Smart contracts based on the rules described above will be built to calculate each individual's votes and output the final decision.

Roadmap:

—MVP: Minimum Constitution provisions for transition between Phase 1 and Phase 2.

—Milestone 1: Basic anti-fraud service for freezing of funds to penalize undesired behavior.

—Milestone 2: Smart contracts implementing voting functionality based on the consensus and decision making rules.

—Milestone 3: Advanced anti-fraud service.

CBR-Finance

On the community economy side, we will develop a self-organizing automatic economic stability system in the form of smart contracts that will mimic the responsibilities and actions of a central bank, mitigating the economic risks around the ecosystem's functioning. There will be four distinct sets of smart contracts - Institution, Taxation, Oracle and Project Development Fund.

Institution

This Institution will regulate the community's finances and economic functioning, including: (1) how tokens will be generated, making valuable contributions to the platform; and, (2) how tokens will be used, tapping into the platform's algorithms. It will set minimum viable transaction limits and a few additional properties - see Section 9 - based on the output from the Oracle. The Institution will also hold reserves of both ETH and AIUR tokens to be able to act aiming at a stable economy. Lastly, in certain situations it will also be able to act as a gas mediator covering some of the gas costs for the transactions when that is seen to be in the best interest of the community.

Oracle

The Oracle is the service that will make external market readings on the AIUR / ETH rate, and provide it for internal use in the smart contracts within project CBR.

Taxation

The taxation service has the goal to serve as a tool for the Institution to stabilize the economy. It will take information from the Oracle, adopt the appropriate tax levels determined by the Institution, and apply the resulting taxes set. The service will have the capabilities of applying taxes both on a per transaction level (when selling of tokens is not beneficial to the community's economy), and also per account (when 'hodling' is counter productive and additional transaction volume is required).

Project Development Fund

The Project Development Fund will be initially established at the end of the crowd sale and will be used to fund the development of all projects outlined in this section. It will act as an escrow smart contract containing 75% of the raised ETH. It will release portions of its funds once a project delivery is released. It will allow contributors to claim ETH from the released amount based on their contribution to the project delivery. The list of concrete project deliveries proposed initially is laid out in the next subsection.

Roadmap:

—MVP: Minimal implementation of the Institution aiming at a functional economy, setting up the reserves, token sale functionality and enforcement of restrictions (i.e. 2% cap). Establishment of the Project Development Fund.

—Milestone 1: Setting up the criteria for escrow release of funds in the product development fund. Setting up the taxation services.

—Milestone 2: Adding basic stability mechanisms to the Institution.

—Milestone 3: Implementation of the Oracle. Adding more advanced stability mechanisms.

5.4 Projected roadmap timeline

In this sub-section we present the envisioned packaging and timeline of the work outlined in the development roadmap.

First, some general guidelines:

—We define project delivery (PD) as meeting a set of milestones that contain major progress, either enabling a new right, a new business functionality for third party applications or a new infrastructural mechanism, for example.

—Each project delivery should be made approximately four months after completion of the previous one. This does not mean that functionalities aimed to be delivered at PD4 should be fully developed in the previous four months. Work could have started much earlier.

—Each PD will be assigned a work-load percentage, and this will mark the portion of funds that will be released when delivery is complete.

—Funds in the Project Development Fund will be allocated to twelve PDs in total.

Full details of milestones and PDs is included in the Appendix and the timeline is visually presented on Figure 3.

For completion, we define PD0 as the token sale setup involving the deployment of initial functionalities (see Figure 4) for the basic operation of the community, which will be covered by Iris.ai and not form part of the Project Development Fund.

The goals for PD1 are:

—To establish the planned economic stability mechanisms that will enable free trading of the AIUR token.

—To introduce the first token reward mechanism - a keyword annotation tool for AI trainers.

The first step in project Char will be needed as a prerequisite for PD1.

PD2 includes finalizing the CBR-Finance project, preparation for community voting (consensus protocol), and set up of project Pylon and project Blackstone (Hypothesis Extraction Engine). It will also introduce the second reward mechanism targeted at coders and bug finders.

The third deliverable (PD3) focuses on project Blackstone and its requirements for further development. It includes the next step in the development of the Hypothesis Extraction Engine, building a hypothesis annotation tool (project Char) and automatic deployment of scripts to ease development and testing (project Pylon). This will bring with it new reward mechanism for hosting services, and deployment of the first APIs for hypothesis extraction as a service to the community, including third party developers and other users.

Rewards for researchers are introduced in PD4, with the MVP from project Alexandria and one-document argument extraction delivered via project Blackstone. Connected new API services will be made available too.

The goals of PD5 will be to initialize the Knowledge Tree Builder (KTB) and the Reproducibility Engine.

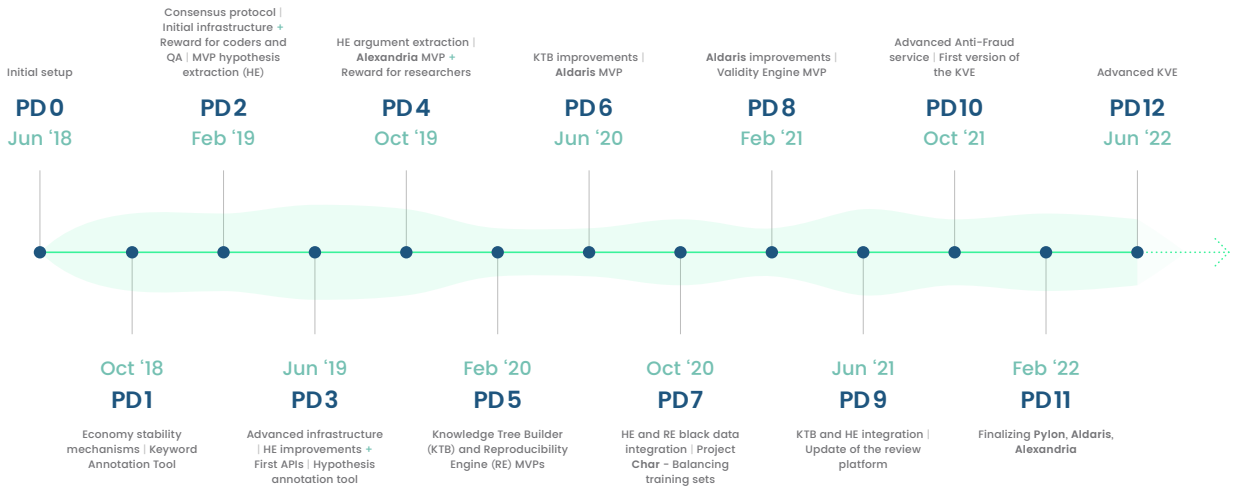

Fig. 3. Projected roadmap timeline

An appeal service for researchers and the next steps in developing the Knowledge Tree Builder are the deliveries included in PD6. This will improve the review platform and introduce new KTB API services.

PD7 focuses on 'black data' integration to both the Hypothesis Extraction Engine (HEE) and the Reproducibility Engine. It also contemplates due improvements to the generation of training sets. This delivery will conclude the APIs planned for the HEE.

PD8 contains the initialization of the Validity Engine, a natural next step in project Blackstone following completion of the HEE. This release will also include the further advancement of project Aldaris.

PD9 is a major delivery. It contains the final version of the KTB, which introduces the possibility of pushing through a major update in project Alexandria's review platform. The problems caused by information overload will be drastically reduced, both for reviewers and rest of the community, using new APIs from the HEE and the KTB.

The goals of PD10 are to conclude the CBR-Governance project, with improvements in the anti-fraud services. Most importantly, it will also include the second milestone of the Validity Engine, and hence the delivery of the first version of the KVE as an API.

PD11 concludes the main goals of the project with the finalization of project Aldaris, project Alexandria, and project Pylon.

Lastly, the final project delivery (PD12) is focused on producing a second, more advanced version of the Knowledge Validation Engine.

5.5 Aiur software license

The terms of the software license governing Aiur's use can be found online at: https://projectaiur.com/software-license/

6. INTRODUCTION TO IRIS.AI

Iris.ai is a seed funded, Singularity University, 500 Startups, Founders Factory, German Technology Entrepreneurship Centre and Creative Destruction Lab backed, three year old international startup developing an AI to democratize access to scientific knowledge. Since founding Iris.ai in 2015 we have focused our efforts on developing a machine that can read and understand scientific text. Our initial trajectory has been covered by Fast Company [Editors 2017], TechCrunch [Kamps 2016], Wired [Gholipour 2017], Science Magazine [Bohannon 2017] or the World Economic Forum [Gaffney and Young 2016], among others.

The Iris.ai tools already in the market have shown valuable results when it comes to aiding researchers explore the research landscape and conduct a literature review through Step 1 (Broad overview of the space) and Step 2 (Narrow down to exact reading list) of the R&D process (see Figure 5). A process that can take up months when done manually can be condensed into a couple of days' work.

This progress has been possible thanks in part to the generous efforts of our over 9,000 strong community of volunteer AI trainers. This community has shown how small individual commitments can have real aggregate impact in improving AI algorithmic quality, constituting a one-of-a-kind case study in network collaboration that we aim to scale and leverage further.

Additionally, Iris.ai has already demonstrated its ability to contribute much needed research advances, hopefully paving the way to future breakthroughs, namely through the peer reviewed publication of our first paper presenting WISDM (Word Importance-based Similarity of Documents Metric)[Botev and Marinov 2017], our in-house researched document similarity metric.

Next in line we want our tool to master Step 3 (Knowledge extraction) and begin tackling Step 4 (Build new knowledge) and Step 5 (Summarize and present). As per some recent views expressed in connection with artificial intelligence advances in the field of radi-

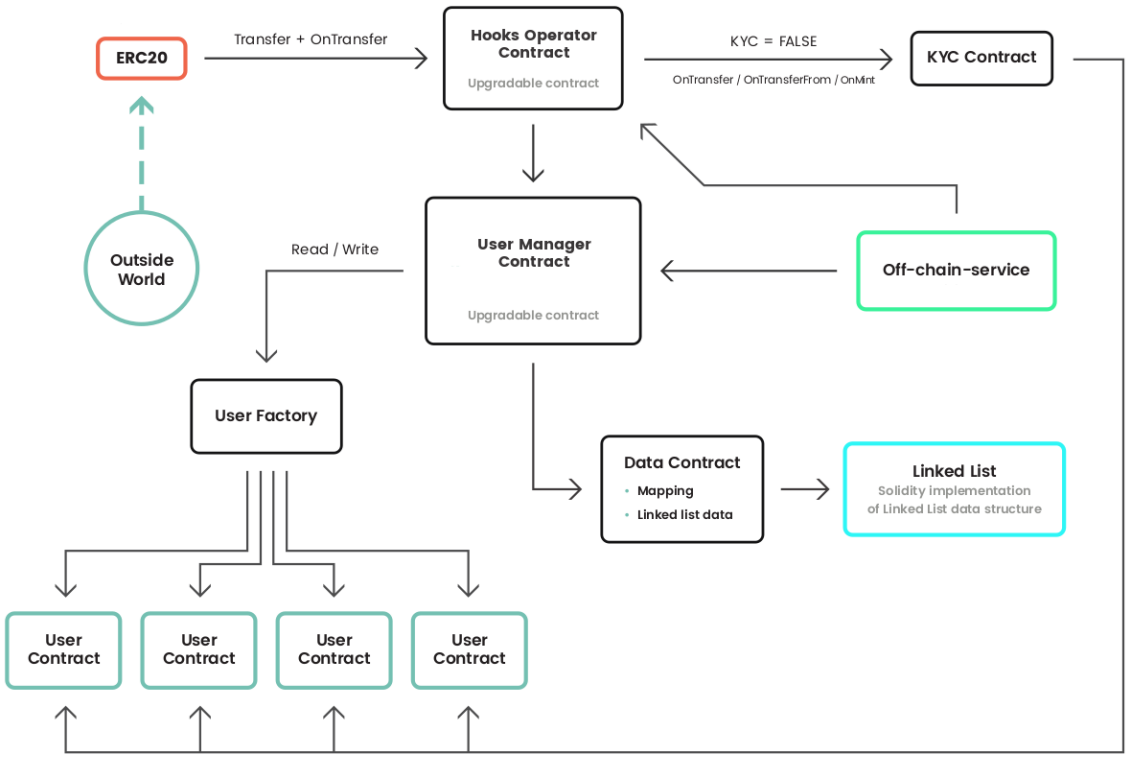

Fig. 4. Smart contracts architecture - PD0

ology [Harvey 2018], we firmly believe that our science assistant software will help human scientists become more effective, easing the pain involved in the drudgery part of their current process. We do not see researchers replaced in any reasonable time frame.

Achieving this development roadmap hinges on the ability to tap into an open Knowledge Validation Engine along the lines described above (see 'The blockchain at play').

7. EXISTING IRIS.AI PRODUCTS

At present Iris.ai solves researchers' time consuming and frustrating workflow-related problem by semi-automating their process. We do this through a combination of product functionalities already available online at https://the.iris.ai.

With the Iris.ai 3.0 version, referred to as the exploration tool, a researcher gives Iris.ai a written problem statement of 300-500 words. Within a few seconds Iris.ai presents the user with a visual overview of the topics of the paper and research papers related to those topics.

What happens in those few seconds is that Iris.ai pulls out the most meaning bearing words in the text, then identifies contextual synonyms from millions of other research papers as well as topic words - hypernyms - and uses these to form a weighted fingerprint of the problem. Then she matches this to the currently connected database using a proprietary document similarity metric (peer reviewed and published). As the tool is already commercially available, we have had the opportunity to validate it.

We have done this through an event format called Scithon - Science Hackathon. A Scithon is an innovative event format developed to help address scientific research challenges in a compressed time frame. In a Scithon groups of interdisciplinary researchers compete using Iris.ai's exploration tool to quickly map out and digest the relevant research around a given challenge. Since it is a competition, a challenge needs to be formulated to structure the event around it. This responsibility is given to an external party, usually an industrial, academic or non-governmental organization (challenge provider). The challenge provider also forms an expert panel that assesses the results and declares the winners. Across all Scithon events run to date participants followed the same process and submitted a standardized report at the end for evaluation. This ensured a fair and transparent procedure for determining the winners.

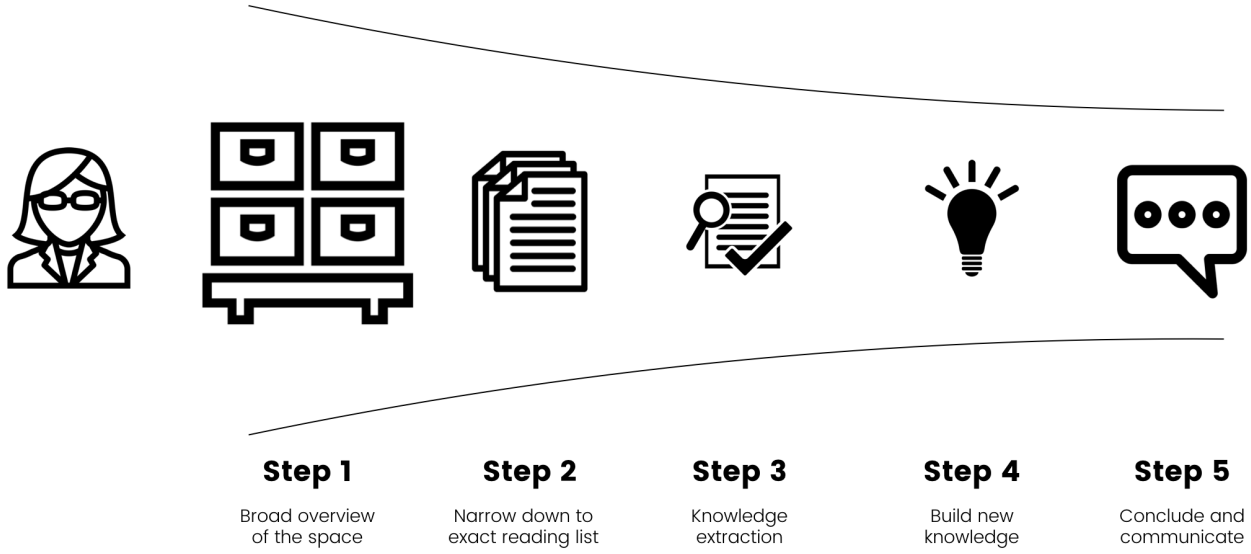

Fig. 5. The researcher's process

The scores provided by the jury for each team together with the key-logger data retrieved from the computers used to centralize each team's research discovery activities were then used to extract the key findings from each Scithon.

A peer-reviewed paper including a cross-sectional overview of the key metrics gathered from the initial series of Scithons run since the launch of the tool in September 2016, laying out all fully comparable data available has been presented at WOSP 2018 and is currently pending publication.

The Iris.ai 4.0 service, also referred to as the systematic mapping study tool, can function as a second step after using our 3.0, the exploration tool, but can also work as a standalone solution.

The starting point is for researchers to have explored the research field around the problem being solved, either via the Iris.ai exploration tool or via an old school keyword query. The researchers in question thus start with a collection of at least 1,000 or as much as c. 20,000 documents, and the goal now is to narrow down the corpus, focusing in on exactly what they are looking for.

This Iris.ai service is a semi-automated process replicating what is in the academic world referred to as a systematic mapping study or a research landscape mapping. Today, this is done without machine help by reading the title of all of the papers and manually marking them for inclusion/exclusion, then diagonally reading all selected abstracts and marking them for inclusion/exclusion, and then fine reading the remaining content. This is obviously an incredibly time consuming process.

The Iris.ai service semi-automates this through an iterative process between the user and the tool, where the user as a Step 1 sets keyword-based inclusion/exclusion criteria from their pre-existing knowledge, and we use formed "fingerprints" (keywords, contextual synonyms and topic modelled words) to select documents for inclusion and exclusion. In Step 2 the Iris.ai service uses Neural Topic Modelling to present the user with the general topics the remaining documents fall into, and the user selects what topics to include and exclude. The model is retrained on the by now limited data set, and the user again selects include/exclude criteria, this time on more finely tuned topics. This is iterated repeatedly until users are left with content falling into exactly the topics they need to read.

This process has been demonstrated in our research efforts together with Swedish Chalmers University of Technology. Tentative results (pending publication) show that a process which today can take three weeks for an industrial researcher, with a self-reported 70% confidence level, can be reduced to less than two days of work (90% time reduction) while upping the confidence level to a measurable 85%.

8. TOKEN SALE OVERVIEW

In this section we present the token sale rationale; our views on the future market for AIUR tokens; the key data around the sale (including targeted amounts, tentative timeline, consideration accepted and financial plan); as well as approaches to distribution strategy, lockup period, escrow mechanism and token pricing.

8.1 Sale rationale

Building Aiur requires devising a mechanism to fund the initial stage of its development suited to support open innovation. Not only this, it also requires incentivising a sufficiently large number of AI Contributors (see Planned ecosystem section) to commit effort at uncertain two-sided value-exchange rates over a long timeline. We believe this can only be reasonably achieved through the creation and issuance of a new functional token running on a trustless, decentralized smart contract framework. With clear 'proof-of-human-work' characteristics in its design, the AIUR token falls within those identified as holding greater value resiliency and po-

tential among the broad scope of functional tokens currently being ideated [Selkis 2018].

As pointed out by Pearson [Pearson 2017], blockchain technology brings with it high levels of coordination with low levels of centralization. On the other hand, annotated data scarcity and training opacity pose a grave threat for the future development of AI, at least in the foreseeable future. Through our initial token sale we intend to play our part in shifting the focus from trust in institutions towards trust in networks, for the open development of natural language processing machine intelligence. The sale of essentially functional tokens offers a unique opportunity to align founders and technologists with researchers, reviewers, trainers, coders and other users, funding the future development of the platform. And we want to achieve that by optimizing for the creation and flow of sustainable social value through a community-owned network, thereby bypassing the limitations of short-terministic profit maximization dynamics.

8.2 A market for AIUR tokens

We have fundamentally modelled the AIUR token as: (1) the sole instrument available for the community to tap into Aiur directly via an Application Programming Interface ('API'); and, at the same time, (2) a potential voucher, i.e. a digital right to purchase products built on top of Aiur at a discount. Shorter term, upon closing of this initial sale, AIUR tokens will be redeemable for services such as Iris.ai premium accounts at a deep discount to Iris.ai's rate card. Far from an instrument suited to short-term financial speculation, AIUR tokens are designed for natural holders, who believe in the value-added that Aiur will bring either to them directly or to other third party use-cases.

As a result, and provided sufficient demand for Aiur-supported products and technology, as initial backers token owners should be able to appropriate the difference in value between token prices and service prices. In addition to this expected source of value, the quality of Aiur's algorithms should improve substantially over time [Elosua 2018], as we have experienced releasing different Iris.ai product versions, from 1.0 to 4.0. We have been able to validate this progress through quantitative mathematical performance tests and qualitative feedback from customers.

It is our firm belief that as development advances we will observe further improvements, which should in turn bring increased value added to a wide user base. And hence, as students, researchers and most importantly, financially strong corporate R&D departments tap Aiur's algorithmic brain and demand its services, we expect the value of the AIUR tokens to increase in parallel. Furthermore, and as noted by Fabrice Grinda [Grinda 2017], we expect decentralized systems will have a very hard time replacing established centralized competitive marketplaces, and that, as a corollary, markets that do not yet exist are more likely to be created on blockchain. We believe AIUR tokens will play a key role in forging a new marketplace for verified, validated knowledge.

Key factors affecting expected market liquidity are discussed below (under Supply and demand policies.)

8.3 Key data

The Aiur token sale will target raising the ETH equivalent of EUR 10,000,000 in overall sale proceeds, with a minimum floor for completion set at 60% and a hard cap of 500%. Individual token purchases will need to be for 5 AIUR or above. No single purchaser will be allocated more that 2% of the total token pool post sale.

As per our current timeline (subject to adjustment), the pre-sale will start on July 10th, 2018, with initial access granted to our community of existing AI Contributors and AI Users and other whitelisted early backers. After this initial pre-sale period, the sale will be opened to the public on September 5th, 2018. the sale will run until the earlier of: (1) the time four times oversubscription relative to the hard cap is reached; or, (2) a four weeks public sale period. Final oversubscription adjusted allocations will be determined in line with the distribution criteria laid out below.

Ether will be the only consideration accepted for orders placed.

We will assign 75% of the amount raised through this initial token sale to fund the ongoing development of the Aiur Knowledge Validation Engine. These funds will be released subject to the achievement of milestones, in an open, competitive environment, subject to community scrutiny and ultimate decision-making. The remaining 25% will be allocated to Iris.ai as payment for the service provision on account of the ideation, design and planning of project Aiur, the formation and initial organization of the community, and to cover other expenses required to successfully kick-off the project. Iris.ai's founders will not receive any direct monetary compensation, in either fiat, crypto currency or AIUR tokens.

8.4 Distribution strategy

Our primary goal distributing AIUR tokens is to align long term incentives with a large and fast growing base of existing community members and early backers, acting as individuals, corporates or institutions. Our secondary goal is to maximize breadth of holders, in particularly among target Aiur users.

For purposes of the initial distribution of tokens, we target long term incentivisation alignment from two angles:

—Firstly, with regards to our existing AI trainer and future AI contributor communities. Since our initial product launch in 2016, we have facilitated the formation of a solid community of AI trainers. This community, currently c. 8,000 strong, has contributed thousands of non-retributed work hours to help us achieve our vision, namely developing a machine to make scientific knowledge more generally accessible. We aim to correct that rewarding initial contributors and project ambassadors with early access to tokens; and,

—Secondly, with regards to our users and clients. To date our tools have been used by approx. 258,000 users, out of which about 21,000 are returning and active users. In addition to that, over the past year we have worked with over a dozen clients. In traditional SaaS pricing models, the corporates and institutions that have led the way trusting us in the development of Iris.ai's algorithmic brain, through initial Scithons, trials and licences, are not set to capture proportionate benefits from the initial impulse provided. They will be granted early access to tokens too.

Besides rewarding existing relationships, the project's marketing efforts will target allocating tokens in advantageous terms to early backers, and in particular both existing and aspiring scientists, via mechanisms including airdrop and/or bounty campaigns. The Aiur token sale will also target maximizing breadth of holders through expansive social media communications.

In terms of AML/KYC requirements, there will be three categories of buyers:

—Anonymous;

—Weakly verified; and,

—Fully verified.

Anonymous accounts with token holdings below the established thresholds (see below) will be permitted in the system. This will

allow for users to express their support of the project in the face of potential threats, which could take various forms (i.e. professional career advancement), particularly in certain jurisdictions.

Anonymous users can not hold in their account more than 60 AIUR tokens at any point in time. They can do transactions of up to 15 AIUR tokens a day, up to 60 AIUR tokens a week, and up to 120 AIUR tokens a month. If they pass the limit they need to do a weak KYC and move to the weakly verified group.

Weakly verified users can not hold more than 280 AIUR tokens at any point in time. They need to provide their names, date of birth and nationality for verification. Then they can do transactions of up to 70 AIUR tokens a day, 280 AIUR tokens a week and 560 AIUR tokens a month.

Fully verified users have no AML/KYC related limits. They need to provide scanned ID, selfie, name, date of birth and nationality for verification. They will also be checked against all available sanction lists.

8.5 Lockup period and escrow mechanism

From the moment the token sale starts and until it completes, consideration received will be provisionally exchanged for AIUR tokens. There will be a lockup period of up to a maximum of two months post sale completion.

During the lock up period additional KYC and AML procedures will be put in place, and Iris.ai will reserve the right to adjust the final distribution of AIUR tokens.

Funds will be returned to token buyers automatically both if the sale does not complete or if there are unfulfilled or scaled back orders. In all fund returns a deduction of up to 3% will be applied to cover estimated direct third party transaction costs.

Post token sale, 75% of the funds raised will be placed in escrow and governed by a smart contract. Funds will only be released upon proof of achieved milestone, following the schedule tentatively outlined in Aiur development roadmap above.

8.6 Token pricing

We have analyzed recent trends applied in previous token offerings. The Ethereum community has experimented with various sale configurations for ERC20 tokens. Capped sales can reach several millions of dollars and sell out in a matter of minutes, leaving buyers unable to participate, disappointed, and frustrated. Uncapped sales, which run without such maximums, provide buyers little clue as to the fraction of total tokens their contribution will ultimately purchase. Other distribution experiments, including hidden caps and reverse Dutch auctions, have suffered similar fates. Indeed increasing purchase power and limited supply may cause buyers in a reverse Dutch auction to jump in too soon [Jason Teutsch and Brown 2017b].

Against this backdrop, we will introduce two straightforward adjustments to our capped token sale: (1) give priority to existing community members and early backers (see Token distribution above); and, (2) limit purchasing orders to a maximum sale cap, with a mechanism to scale orders back to enforce the 2% cap relative to the total number of tokens post sale.

AIUR tokens will be sold to purchasers at consistent prices, as a factor of time-to-closing. We will divide the sale into weekly tranches and introduce stepped price increases for each successive tranche. Public sale discounts will range from 20% to 0%.

Furthermore, we do not intend to incentivize any third party to support our token sale through the opaque issuance of tokens at no cost or deeply discounted rates. Any such awards and the agreements covering them will be fully transparent before the public token sale kicks off.

AIUR tokens will be priced at ETH 0.01 per token (pre discount). The total number of tokens post sale will be the result of adding: (1) the tokens sold to raise the ETH equivalent of EUR 6,000,000 (minimum floor) to EUR 50,000,000 (hard cap), as per the ETH/EUR exchange rate set at the start of the pre-sale, factoring in discounts; (2) the tokens issued through promotional airdrop and/or bounty campaigns; (3) the tokens minted to fund the Institution's AIUR reserves; and, (4) an amount of tokens issued to Iris.ai to guarantee Phase 1 governance stability (i.e. 50% plus one of the votes).

9. SUPPLY AND DEMAND POLICIES

In this section we examine the project's nature and its implications on token holders; envisaged use-cases buying and selling AIUR tokens; the key fiscal and monetary policy mechanics in place; and potential issues that might arise with their respective mitigants, all in connection with the project's anticipated supply and demand dynamics.

9.1 Project nature and implications

Developing Aiur is a multi-year project running until 2021 in our initial estimates, around which a sufficiently large community of active core users and contributors needs to rally long term.

This means that, in the road towards building that community: (1) there is a requisite to provide relative stability throughout the time horizon required to develop the tool; (2) token uses and token contributions need to surpass certain volume thresholds; (3) token holders are not to hodl tokens and should largely overlap with targeted token users; and, (4) token uses should be affordable to aspiring users.

With this in mind, we have approached token holder rights from a seemingly restrictive viewpoint. All of the project's governance provisions are aimed at building a viable long-term ecosystem around a functional tool. Thus, acquired and/or generated AIUR tokens will not grant holders any individual rights found to be in conflict with Aiur's mission, as declared by the project's governance bodies in the respective phase.

Enshrining Aiur's mission implies several things, including for example an acceptance of potential absence of financial returns. It also includes expressly stating in the project's communications that declining token prices relative to other measures of value will not be considered, in isolation, as a negative key performance indicator.

9.2 Envisaged use-cases

AIUR tokens will be generated every time an Aiur contributor submits an accepted contribution to the system, for instance, when an individual AI trainer submits a successful training input. Once an annotation quality algorithm (initially Iris.ai-developed) verifies the input received from the submitting AI trainer, generating a transparent, individual score per training input submitted, a smart contract will be triggered to register the token award. In the near future, and as a next step, we plan the introduction of additional token generation mechanisms that are congruent with Aiur's product roadmap, i.e. through accepted code commits (see 'Aiur development roadmap' section).

Let's consider the initial scenario whereby a successfully trained input, once transparently verified, will yield the submitting AI trainer one token: 1 successful input = 1 token awarded, setting the initial token award rate at 1. Upon successful completion of the initial token sale, two things will happen. Firstly, successful

AI contributor inputs will yield tokens at a token-award rate of 1. Secondly, Iris.ai's product offering will be tokenized, with AIUR tokens enabled as a significantly discounted payment mechanisms for all of Iris.ai's fiat denominated products and services. Furthermore, in addition to the tokenization of Iris.ai's existing product offering, our current and future customers will need to acquire AIUR tokens in order to tap into the Aiur Knowledge Validation Engine via API, initially at a token-use rate of 1, meaning: 1 API call = 1 token used.

As per the Aiur development roadmap section, the Aiur Knowledge Validation Engine will comprise several sub-systems. There will be a base product price level for an API call, with the price of each service set by reference to this base price.

Beyond the initial token sale, we expect customer use-cases driving token demand to come in several forms: (1) a researcher (individual) writing scripts herself without a software client; (2) a corporate (team) needs to build their own client or hires a third party to do that for them, possibly deploying a private instance; and, (3) an app developer creating an entire new tool (a machine), and repackaging services as an intermediary for a freshly targeted customer base.

Importantly, after the envisaged initial token generation event Iris.ai will become both: (1) a committed, high volume AI User, driving AIUR token demand; and, (2) a significant AI Contributor, with its development team committing code to improve the Aiur Knowledge Validation Engine, generating tokens in the process.

In connection with (1) above, as the company continues implementing its product development roadmap, the present and future suite of Iris.ai tools will tap into the Aiur Knowledge Validation Engine as the preferred source of validated input data. In relation with (2) above, the development of software products on top of the Knowledge Validation Engine should provide an intensive, client-driven use-case contributing recognized value to the platform.

This conceptual architecture requires implementation of an independent, community-driven platform governance system to guarantee fair market dynamics and prevent one-sided abuse. We have proposed some ground rules to collectively generate such a system over time (see 'Constitutional provisions' above).

9.3 Key policy mechanics

As covered before, a functional Knowledge Validation Engine demands broadly balanced, above minimum thresholds, demand and supply flows of AIUR tokens. Significant, sustained mismatches would jeopardize the achievement of the project's long-term goal, namely, democratizing science through, among others, a realignment of agent incentives.

We intend to achieve this desired equilibrium through: (1) a smart contract-enabled Constitution setting the ground rules, with a commitment, for example, to capping new token issuance at proven value contributed into Aiur; and, (2) the implementation of a smart contract-based Oracle to regulate the market actions undertaken both monetary and fiscal on the back of continuous monitoring of the project's token pool evolution.

Identified participants:

—Buyers - participants in the system that want to buy tokens.

—Sellers - participants in the system that want to sell tokens.

—Producers - contributors to the community with their own expertise (researcher, coder, QA, AI trainer).

—Suppliers - third party developers that create applications to the project APIs (or more generally users of the API).

—Users - users of applications, products or tools build on top of the project's APIs.

—Institution - collection of smart contracts that operate in order to keep the community operational.

The Institution will contain mechanism such as:

—Oracle - service that gathers information from the outside world of the current rate between AIUR token and ETH. Based on that rate it should be able to set a minimum viable transaction limit 'Tmin' and operate in the market transacting tokens. It will be useful for some of the scenarios described below.

—AIUR token reserve - community account that contains AIUR.

—ETH token reserve - community account that contains ETH.

—Gas Mediator - service that should pay the gas for certain transactions

—Tax Man - service that based on the rate between AIUR and ETH computes a tax level and applies a tax to all AIUR transactions. The tax is calculated based on the status of seller. Tax level should be a function of four factors:
  —How much of the stake of the seller are generated tokens vs. acquired ones.
  —How long has the seller held the tokens.
  —Is the account of the seller public or anonymous.
  —Is this a transaction where the Institution is involved.

The Institution is set to regulate the consensus mechanisms in effect to define the operation of the community:

—Constitution - definition of the governance mechanisms plus consensus mechanisms.

—Laws - set of rules defining the rights, obligations of the members, plus measures to be taken if those are violated.

Aside from its regulatory function, the Institution will also operate as an economic actor (monetary and fiscal), including the proactive initiation of market activities buying or selling AIUR tokens on behalf of the community.

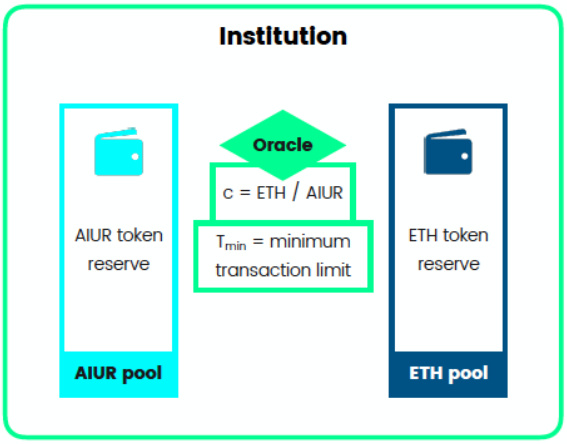

Fig. 6. Aiur supply and demand mechanics (1/6)

As reflected in Figure 6, the project's token pool will initially hold two assets: (1) reserves of the token sale's ETH proceeds, in a sufficient, yet to be quantified amount; and, (2) a market value

equivalent unissued amount of AIUR tokens. As token generation and token use begin to take place, the token pool's Oracle will sample various indicators to track the evolution of relative prices. On the back of this continued market reads, the Oracle will operate in the market with the goal of providing stability to the system.

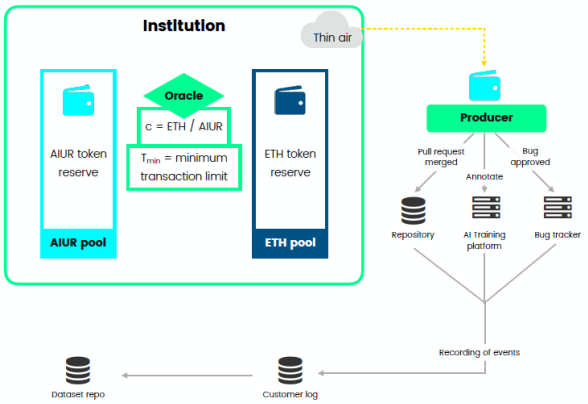

Fig. 7. Aiur supply and demand mechanics (2/6)

In Figure 7, we consider a token generator who submits a new valuable contribution to Aiur. This submission kick-starts a largely automated verification process, which after successful completion will result in the set volume of AIUR tokens being issued to the contributor. These contribution and token generation will be registered in a private blockchain, increasing the volume of tokens in circulation. As tokens generated are subsequently transacted, with a corresponding entry in the Ethereum blockchain, intermediation ETH gas costs would be borne by the project's token pool, hence depleting the pool's ETH reserves.

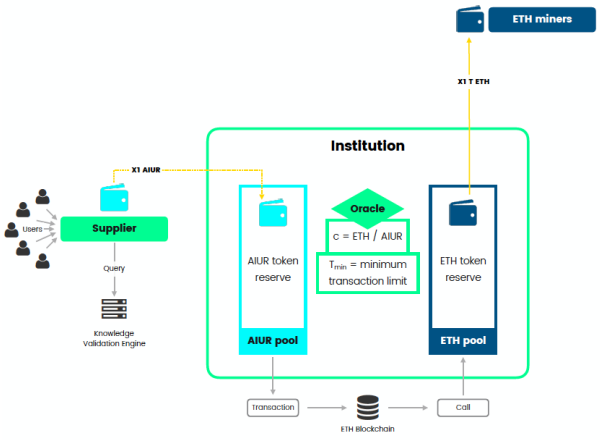

Fig. 8. Aiur supply and demand mechanics (3/6)

As per Figure 8, every time a token user, acting as an end user or through a third party built intermediate product, queries Aiur's algorithmic brain the project's AIUR token pool increases, but intermediation ETH gas costs would again be borne by the project's token pool. Therefore, token uses would increase the stock of AIUR held by the pool whilst depleting the pool's ETH holdings.

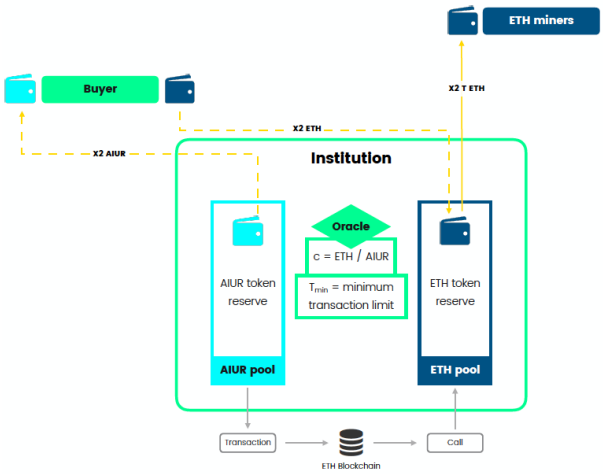

Fig. 9. Aiur supply and demand mechanics (4/6)

Figure 9 reflects open market transactions where the Institution is the counter party transacting with a willing buyer or seller. As programmed through a set of smart contracts, the pool's Oracle has the mission of aggregating token generation and token use flows to execute buying and selling transactions aimed at preserving the token pool reserves at stable levels, relative to each other, adjusted for changing market prices.

Unlike in other token sale processes, this token-generation architecture means that our token supply will hence be variable. Going forward, in well-defined windows after each version release, the project's governing entity will assess: (1) customer demand for AIUR tokens from AI Users; and, (2) the token generation rate of AI Contributors. With this information and the token price observed evolution in mind, Iris.ai will adjust the following four parameters:

—Supply side: Token-award rate.

—Supply side: Supplemental volume of tokens required.

—Demand-side: Token-use rate.

—Demand-side: Product price token exchange rate.

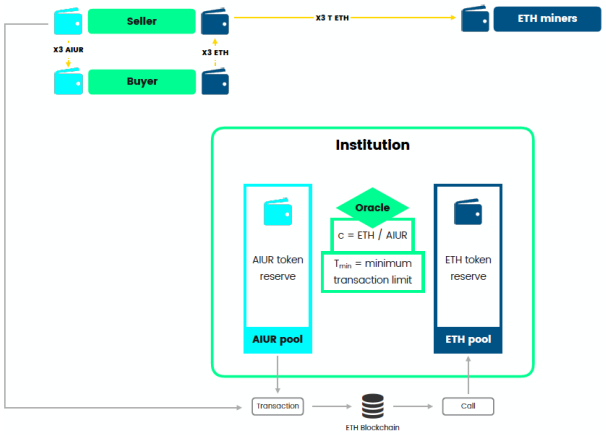

Fig. 10. Aiur supply and demand mechanics (5/6)

Figure 10 shows a transaction between a willing seller and a third party buyer of AIUR tokens. Miner costs to annotate the transaction are assumed to be borne by the seller, without a direct impact in the pool's ETH and AIUR reserves.

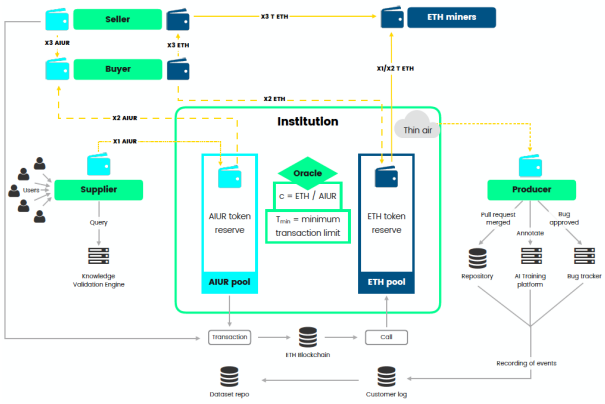

Fig. 11. Aiur supply and demand mechanics (6/6)

In Figure 11 we can see the full mechanism at play, aggregating all the different use-cases described above.

9.4 Potential issues and mitigants

Let's recap. We have designed several layers in the project to create a solid, long-term focused architecture aimed at attaining a minimum, indispensable level of relative short-term, lower volatility token stability. Firstly, through smart contracts establishing Constitution and Institution mechanisms. In line with Ethereum's design, our intent is that token supply will increase following transparent, well-defined rules enabled by one or more smart contracts. This primary layer sets the course limiting discretionary AIUR token supply that could fuel inflation (i.e. a purchasing power decrease in individual token values), and enabling a mitigating reaction to market signs revealing imbalances.

As a secondary layer aimed at securing relative stability, we have designed additional mechanisms for the tentative neutralization and/or mitigation of eventual persistent imbalances, to be resorted to during the project's life, be it during Phase 1 by Iris.ai as initial promoters or during Phase 2 by the Aiur self-governing community . These mechanisms can be best understood approaching them under different scenarios.

Project-specific token appreciation risk. In this scenario, where AIUR token demand greatly exceeds supply, and as per the Constitution provision, the project's governing body will be prompted to review the relative token-generation and token-use rates, and/or approve the release of a new, well-defined, token generation program (i.e. gather an additional set of annotations for algorithmic training purposes, with an X limit on total number of papers trained and Y maximum annotations per paper).

Project-specific token depreciation risk. In this scenario, where token demand lags behind token supply, the additional mechanism, beyond reviewing the relative token-generation and token-use rates, consists of penalizing passive AIUR token balances (i.e. hodler profiles) applying gradual negative interest rates to incentivize higher tool demand and hence increased token circulation. This should result in the Oracle's ability to restore depleted token reserve balances through open market operations.

ETH decoupling risk. In this scenario the intermediation cost (i.e. mining fees) of registering token generation and/or token use activity in the Ethereum blockchain becomes anti-economic. The mechanism contemplates an Oracle-induced token policy change, aggregating transactions in larger blocks for registration. This implies the creation of shadow, non-blockchain logged, notional transaction records.

In terms of factors affecting taxes applied to the sellers in AIUR token transactions, they will include:

—Short-term market fluctuations.

—Potential sustained system instability.

—The perceived merit of the transaction from a community viewpoint.

This assessment of the perceived merit of each token sale transaction will include the evaluation of criteria such as:

—Token generator vs. purchaser.

—Public vs. anonymous identity.

—Short vs. long term holding period.

Innovating in how blockchain projects set up their community mechanics, and in consonance with Aiur's project governance provisions (see Project nature and implications), transaction taxation will ensure that no individual rights found to be in conflict with Aiur's mission will have prevalence over the shared community vision. These rules over token taxation, as well as the rest of the project's governance provisions, are aimed at building a viable long-term ecosystem around a functional Knowledge Validation Engine.

10. ADDITIONAL DATA

10.1 Team

Fig. 12. The Iris.ai team in Berlin

Our team is currently sixteen member strong, combining proven research, development, business and operations expertise.

Anita Schjøll Brede, CEO. Anita leads the Iris.ai team, in areas from strategy formulation and communicating the company's

vision to talent attraction and retention. Anita has a Master's degree in Entrepreneurship and Business design from Chalmers University of Technology, and has had stints at six different universities including Stanford and Berkeley. Before co-founding Iris.ai at Singularity University (GSP '15), Anita was already a serial entrepreneur with four previous start-ups under her belt, in Silicon Valley, Sweden and Norway. Anita has been selected as one of Inspiring Fifty Nordic's most Inspiring women in tech. Twice a TEDx speaker, she has lead Iris.ai as a 500 startups, Founders Factory, SU Global Grand Challenges Awards and TechCrunch Disrupt Startup Battlefield alumni.

Victor Botev, CTO. Victor is responsible for the technology and R&D team at Iris.ai. Victor earned two individual Master's degrees in Artificial Intelligence and Computer Systems and Networks from Sofia University St. Kliment Ohridski and Chalmers University of Technology, respectively. After finalizing his degrees and before co-founding Iris-ai in 2015, he worked at at Chalmers University of Technology, where he conducted research on clustering and predictive neural networks models and the usage of signal processing techniques in studying Big Data. At Iris.ai, Victor also leads overall product development for both the AI and Blockchain workstreams. The team's current R&D efforts focus on unsupervised machine learning for hierarchical concept extraction as well as defining and executing on project Aiur's development roadmap.

Maria Ritola, CMO. Maria leads marketing, sales and business development at Iris.ai, including working hand in hand with both university and corporate clients. Maria has a Master's degree in economics from Helsinki School of Economics. Prior to co-founding Iris.ai at Singularity University (GSP '15), Maria worked as a researcher and a Vice Chairman of the Board at the Nordic Think Tank Demos Helsinki and as an economist at Bank of Finland. Maria has published joint research papers e.g. in MIT Press Journal and Futures. Maria has also worked at UNICEF HQ and co-founded Peloton Club, a leading start-up accelerator focused on CleanTech. Maria has been selected among the most influential women in tech in Scandinavia and the 2017 young person of the year in Finland.

Jacobo Elosua, CFO. Jacobo is responsible for running the financial management, business operations and corporate development of Iris.ai. Jacobo holds three Master's degrees, in International Business (ICADE, E-4), Law and Economics (UNED). Before co-founding Iris.ai at Singularity University (GSP '15), he served nearly 10 years at UBS Investment Bank, where he acted as Executive Director specializing in Media and Technology M&A and financing. In his trajectory as an entrepreneur, he co-founded Ezaro Media, an advisory and investment boutique, and i.ngen.io, a data visualization start-up, among other projects. Jacobo has been very active in the Open Data and Open Government spaces for over a decade. He is also a co-founder and currently acts as Chairman of Civio, a non-for-profit fostering an active and engaged citizenship through transparency, technology and journalism.

Georgi Dimitrov, Head of Development. Georgi leads Iris.ai's software development efforts. Georgi has a Bachelor's degree in Software Engineering and a Master's degree in Technological Entrepreneurship from Sofia University St. Kliment Ohridski. His professional experience includes blockchain technologies, frontend development (SAP UI 5, HTML5, JSF, Java Swing), backend development (JPA, JDBD, JavaSE, WS/SOA and RMI), databases (SQL, SAP HANA, MaxDB, MySQL), quality engineering (Selenium, Mock frameworks and Java test frameworks) security activities (Scanning, Auditing, Black Box & White Box security testing, Penetration testing), web development, software architecture design, software development and quality management. Before joining Iris.ai, he worked for almost 10 years at the German software multinational SAP (SAP Labs Bulgaria), where he was appointed expert software developer and product manager, among other titles.

Ami (Anne-Marie) Defesche, Head of Community. Ami runs and amplifies Iris.ai's community efforts. Ami has a Bachelor's degree in Graphic Design from St. Edward's University. Prior to joining Iris.ai, she was a Senior Community Manager at Blackbaud, where she managed multiple online communities under the Digital Marketing vertical. Previous positions also include being a Community Manager at SpaceTime, a Lead Designer/Developer at Go9Media and a Senior Game advisor at Gamestop. At Iris.ai Ami helps design and executes the company's global community engagement strategy. She also monitors and analyses community statistics and trends, and fosters best-in-class support, engagement and user experience.

In addition to the members listed above, at present the team also includes Volodymyr Krekhovetskyi, Eugene Gurikov, Igor Zhun, Karita Kasurinen, Ronin Wu, Valentin Stauber, Iryna Belotserkovets, Stefan Vasilev, Rosen Martev and Antonia Chekrakchieva.

The project's blockchain engineering efforts combine our in-house development team headquartered in Sofia, Bulgaria, with the consultancy services provided by LimeChain, one of the more prestigious blockchain development teams in operation since 2013.

We have retained the services of ChainSecurity AG, a leading provider developing the first crypto contract automated formal audit platform, for the technical review of project Aiur's smart contracts.

10.2 Advisors and acknowledgements

We are grateful to have an amazing group of advisors to project Aiur. In no particular order they currently include:

—**PD Md. Sönke Bartling**, associated researcher at Alexander von Humboldt Institute for Internet and Society. His focus is the blockchain revolution and what it could mean for science and knowledge creation.

—**Joeran Beel**, Assistant Professor in Intelligent Systems at Trinity College Dublin. His work focuses on machine learning, text mining, natural language processing, the blockchain and other technologies, in areas including recommender systems, search engines, news analysis, plagiarism detection, and machine translation.

—**Peter Suber** is a philosopher specializing in the philosophy of law and open access to knowledge. He is a Senior Researcher at the Berkman Klein Center for Internet & Society, Director of the Harvard Office for Scholarly Communication, and Director of the Harvard Open Access Project (HOAP). Suber is known as a leading voice in the open access movement.

—**Christian Berger** is Associate Professor and Docent for Software Engineering at the Department of Computer Science and Engineering at University of Gothenburg, Sweden. His research focuses on systematically architecting complex software and systems embracing continuous integration (CI), continuous deployment (CD), and continuous experimentation (CE) for a growingly automated and digitalized society.

—**Christoper Fabian** is the former Advisor to the Office of the Secretary General of the United Nations on Data Science as well as an open source venture capital investor and advisor on innovation in international development.

—**Pascal Finette**, Chair for Entrepreneurship and Open Innovation at Singularity University. Formerly Director of Open Innovation

at Mozilla, and author of The Heretic - Leadership in exponential times.

—**Jamer Hunt** is the Vice Provost for Transdisciplinary Initiatives at The New School, where he was a founding director (2009-2015) of the graduate program in Transdisciplinary Design at Parsons School of Design. He has published over twenty articles on the poetics and politics of design.

—**Petr Knoth**, Senior Research Fellow in Text and Data Mining at the Knowledge Media Institute and the founder of CORE, a service that aggregates millions of open access articles from around the world. His research is focused on the domains of text-mining, digital libraries and open access/science.

—**Kent Langley**, Chief Scientist at the ExOFoundation. He has primary accountability for research and implementation of blockchain technologies in the global Exponential Organizations (”ExO”) Community including the CivX Economy & Blockchain. He is a deep technologist and builder of distributed systems, building such systems from the ground up at scale for 20 years.

—**Chris Matys**, founder of Ignite AI, Chief Analytics Officer at Georgian Partners and member of Creative Destruction Lab. Always focused on the impact side, Chris has deep expertise in applied analytics and artificial intelligence, and has developed frameworks to accelerate and scale growth-stage companies post-investment.

—**Ramez Naam**, an American professional technologist, Singularity University faculty member and writer. He is best known as the author of the Nexus Trilogy. He has also authored ”More than Human: Embracing the Promise of Biological Enhancement”. Broadway Books, 2005, and ”The Infinite Resource: The Power of Ideas on a Finite Planet”. University Press of New England, 2013.

—**Amit Pradhan**, founder and President of the Silicon Valley Blockchain Society, founder of Zero AI and partner at JetVentures. He is a seasoned startup founder, investor and strategist operating in the intersection of blockchain and AI.

—**Jon Tennant** is a paleontologist, founder of the preprint server paleorXiv, founder of the Open Science MOOC, and also leading the Foundations for Open Scholarship Strategy Development. He is also the lead author of published papers on both the academic, societal and economic case for Open Access and on the future of peer review.

—**Tamara Giltsoff** is Head of Innovation at the Department for International Development (DFID) at UK Aid. Tamara is a design thinker, impact entrepreneur, tech pioneer and purpose-led leader with a special focus on emerging technologies in developing countries.

In addition to the advisors above, the thoughts, ideas, discussions and concepts culminating in this white paper - and more importantly, the project proposed in it - would not have come to fruition had it not been for a long row of people helping and supporting us along the way.

It is daunting and practically impossible to mention them all but we like impossible challenges and so we would like to personally thank Lasse Birk Olesen, Ethan Buchman, Dennis Benny, Louis Warner, Elizabeth Hunker, Shadi Al’lababidi Paterson, Jørgen Bø, Paul Willinsky, Sebastial Riedel, Nicolai Wadstrom - and if you are not mentioned but feel you should have been, you are probably right!

We would also like to thank Iris.ai’s early seed investors: Tharald Nustad, Anders Lier, Lise Reichsteiner and Nordic Impact, Sjur Dagestad and Sjur Thorsheim, Bjarne Melbye of 2D2M, Anne Worsøe of Bakken & Baeck, INDEX: Design to improve life, Planet 9 Capital, Øyvind Stordalen, Thomas Berglund, Philipp Haydn and Sean Percival, as well as the entire team at Singularity University, 500 Startups, Founders Factory and GTEC Labs.

An finally, an eternal thank you for your patience with us goes to Noelle, Iva, Antti and Ole. We love you.

10.3 Legal and regulatory

The sale of AIUR tokens will be conducted through our EU-based whole owned development subsidiary, Iris.ai BG EOOD. Iris.ai BG EOOD is a limited liability company incorporated under the laws of Bulgaria.

This white paper has been issued during the development of Iris.ai BG EOOD’s project Aiur as an indicative document, and is subject to discretionary future version changes. Any part thereof and any copy thereof must not be taken or transmitted to any country where distribution or dissemination of token sales or initial coin offerings, like the one described in this white paper, is prohibited or restricted.

AIUR tokens are not intended to constitute securities in any jurisdiction, and as such they do not give any rights to dividends or interest. AIUR tokens are not shares and do not give any right to participate to the General meetings of Iris.ai BG EOOD. AIUR tokens are not a digital currency, commodity, or any other kind of financial instrument and they have not been registered under the securities laws of any country, including the securities laws of any jurisdiction in which a potential token holder is a resident.

AIUR tokens are utility tokens and cannot perform or have a particular value outside the Aiur business platform. Therefore, this white paper cannot constitute a prospectus or an offer document for investment in securities. AIUR tokens shall not be used or purchased for speculative or investment purposes. Laws and acts that ensure disclosure and represent regulatory scrutiny for investors’ protection are not applicable in this case. Every buyer of AIUR tokens should look for proper advice in order to understand whether the purchase of the token is appropriate for them or not. The buyer of AIUR tokens undertakes that she/he understands and has significant experience in crypto currencies, blockchain systems and services, and that she/he fully understands the risks associated with the crowd-sale as well as the mechanism related to the use of crypto currencies (incl. storage of such). Iris.ai BG EOOD shall not be responsible for any loss of tokens or situations making it impossible to access the token and/or the services built as part of project Aiur, which may result from any actions or omissions of the user or any person undertaking to acquire AIUR tokens, as well as in case of a hacker attack.

This white paper does not constitute or form part of any opinion on any advice to sell, or any solicitation of any offer by Iris.ai BG EOOD to purchase AIUR tokens, or provide help in an investment decision. Any information in the white paper is provided for general information purposes only and Iris.ai BG EOOD does not provide warranty as to the accuracy and completeness of this information.

Iris.ai BG EOOD is not a financial intermediary according to Bulgarian law and general EU regulations, and as such the company is not required to obtain any authorization for anti money laundering purposes. Additionally, as seller of utility tokens anti money laundering regulations are not applicable, insofar as the main reason for issuing the tokens is to provide access rights to a non-financial application of blockchain technology. However, good practices require the verification of buyers’ identity. Also, more and

more European authorities are discussing a future regulation of the token sale and initial coin offering processes, which will also include explicit KYC and AML provisions. It is highly possible that soon AML obliged entities might be broadened to include companies performing initial coin offerings and token sales. The KYC process is the only way Iris.ai BG EOOD can check the source of funds raised during the token sale, and Iris.ai BG EOOD will aim to do so by verifying buyers identities above a predefined buying threshold. Explicit checks and verification of sanction lists may also apply.

Buyers are not eligible and are not permitted to participate in the AIUR token sale (as referred in this white paper) if they are citizens or legal entities, resident or incorporated (with address, tax or otherwise) or green card holders of the USA or a resident of the People's Republic of China, South Korea or Iran. The same pertains for residents of the Republic of Singapore, Canada and Australia. Such Restricted Persons refer to any firm, company, partnership, trust, corporation, entity, government, state or agency of a state or any other incorporated or unincorporated body or association, association or partnership (whether or not having separate legal personality) that is established and/or lawfully existing under the laws of the restricted jurisdiction, listed above.

The sale of tokens has not been registered under the U.S. securities act of 1933, as amended (the securities act), or under the securities laws of certain states. Tokens may not be offered, sold or otherwise transferred, pledged or hypothecated except as permitted under the act and applicable state securities laws pursuant to an effective registration statement or an exemption there from.

Acquiring AIUR tokens shall not grant any right or influence over Iris.ai BG EOOD's organization and governance to the buyers.

Regulatory authorities are carefully scrutinizing businesses and operations associated to crypto currencies and initial coin offerings in the European Union. In that respect, regulatory measures, investigations or actions may impact Iris.ai BG EOOD's business and even limit or prevent it from developing its operations in the future. Any person or legal entity undertaking to acquire AIUR tokens must be aware of Iris.ai BG EOOD's business model. Additionally, this white paper may change significantly or need to be modified because of new regulatory and compliance requirements from any applicable laws in any jurisdictions. In such a case, purchasers and anyone undertaking to acquire AIUR tokens acknowledges and understands that neither Iris.ai BG EOOD nor any of its affiliates shall be held liable for any direct or indirect loss or damage caused by such changes.

Iris.ai BG EOOD will do its utmost to launch its operations and develop the project Aiur platform, including the intended Engine for Knowledge Validation. Anyone undertaking to acquire AIUR tokens acknowledges and understands that Iris.ai BG EOOD does not provide any guarantee that it will manage to achieve it. They acknowledge and understand therefore that Iris.ai BG EOOD (incl. its management bodies and employees) assumes no liability or responsibility for any failure or downfall that would result from or relate to the incapacity to use AIUR tokens, except in case of intentional misconduct or gross negligence.

No EU regulatory authority has examined or approved any of the information set out in this white paper. No such action has been or will be taken under the laws, regulatory requirements or rules of the European Union.

10.4 Contact

For additional information, you can visit us at https://iris.ai, join our Telegram channel at https://t.me/theirisai, follow us in social media or write to founders@iris.ai.

REFERENCES

Neil Malhotra Annie Franco and Gabor Simonovits. 2014. Publication bias in the social sciences: Unlocking the file drawer. *Science* (2014). http://science.sciencemag.org/content/345/6203/1502.long.

John Bohannon. 2017. The cyberscientist. *AI Transforms Science - Science Magazine* (2017). http://science.sciencemag.org/content/357/6346/18/tab-figures-data.

Victor Botev and Kaloyan Marinov. 2017. Word Importance-based Similarity of Documents Metric (WISDM). *ACM Digital Library* (2017). https://dl.acm.org/citation.cfm?id=3127530.

Stephen Buranyi. 2017. Is the staggeringly profitable business of scientific publishing bad for science? *The Guardian* (2017). https://www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science.

Vitalik Buterin. 2017a. N/A. *Twitter* (2017). https://twitter.com/VitalikButerin/status/945987507941978112.

Vitalik Buterin. 2017b. Notes on Blockchain Governance. *Vitalik Buterin's website* (2017). http://vitalik.ca/general/2017/12/17/voting.html.

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science magazine* (2015).

David Dinkins. 2017. ICO to Build Next Generation AI Raises $36 Million in 60 Seconds. *Cointelegraph* (2017). https://cointelegraph.com/news/ico-to-build-next-generation-ai-raises-36-million-in-60-seconds.

Editors. 2017. The 2017 World's Most Innovative Companies. *FastCompany* (2017). https://www.fastcompany.com/most-innovative-companies/2017/sectors/artificial-intelligence-machine-learning.

Editors. 2018. COREs Open Access content has reached the Moon! (or how about them stats 2017 edition). *CORE's blog* (2018). https://blog.core.ac.uk/2017/12/21/cores-open-access-content-has-reached-the-moon-or-how-about-them-stats-2017-edition/.

Fred Ehrsam. 2017. Blockchain Governance: Programming Our Future. *Medium* (2017). https://medium.com/@FEhrsam/blockchain-governance-programming-our-future-c3bfe30f2d74.

Jacobo Elosua. 2018. Not all utility tokens are created equal. *Medium* (2018). https://medium.com/@elosuaj/not-all-utility-tokens-are-created-equal-d7abf18fc1c8.

Jon Evans. 2017. Opinion: After the end of the startup era. *TechCrunch* (2017). https://techcrunch.com/2017/10/22/ask-not-for-whom-the-deadpool-tolls/.

Rose Eveleth. 2014. Academics Write Papers Arguing Over How Many People Read (And Cite) Their Papers. *Smithsonian Magazine* (2014). https://www.smithsonianmag.com/smart-news/half-academic-studies-are-never-read-more-three-people-180950222/.

Bernard Forgues and Sbastien Liarte. 2013. Academic Publishing: Past and Future. *M@n@gement* (2013). https://www.cairn.info/revue-management-2013-5-page-739.htm.

Owen Gaffney and Denise Young. 2016. 5 ways artificial intelligence will disrupt science. *World Economic Forum* (2016). https://www.weforum.org/agenda/2016/06/5-ways-artificial-intelligence-will-disrupt-science/.

Bahar Gholipour. 2017. Artificial intelligence could dig up cures buried online. *Wired Magazine* (2017). https://www.wired.com/2016/11/artificial-intelligence-dig-cures-buried-online/.

Fabrice Grinda. 2017. Some thoughts on cryptocurrencies. *Fabrice Grinda's blog* (2017). https://fabricegrinda.com/some-thoughts-on-cryptocurrencies/.

Hugh Harvey. 2018. Why AI will not replace radiologists. *Medium* (2018). https://towardsdatascience.com/why-ai-will-not-replace-radiologists-c7736f2c7d80.

Matthew Hutson. 2018. Missing data hinder replication of artificial intelligence studies. *Science* (2018). http://www.sciencemag.org/news/2018/02/missing-data-hinder-replication-artificial-intelligence-studies.

Vitalik Buterin Jason Teutsch and Christopher Brown. 2017a. Interactive coin offerings. (2017). https://people.cs.uchicago.edu/ teutsch/papers/ico.pdf.

Vitalik Butarin Jason Teutsch and Christopher Brown. 2017b. Interactive coin offerings. *N/A* (2017). https://people.cs.uchicago.edu/ teutsch/papers/ico.pdf.

Arif E. Jinha. 2010. Article 50 million: an estimate of the number of scholarly articles in existence. *Learned Publishing* (2010). https://doi.org/10.1087/20100308.

Steven Johnson. 2018. Beyond the Bitcoin Bubble. *Ney York Times* (2018). https://www.nytimes.com/2018/01/16/magazine/beyond-the-bitcoin-bubble.html.

Brad Plumer Julia Belluz and Brian Resnick. 2016. The 7 biggest problems facing science, according to 270 scientists. *Vox* (2016). https://www.vox.com/2016/7/14/12016710/science-challeges-research-funding-peer-review-process.

Haje Jan Kamps. 2016. Iris AI drastically expedites research through the power of artificial intelligence. *TechCrunch* (2016). https://techcrunch.com/2016/10/25/iris-ai-for-science/.

Adam Ludwin. 2017. A Letter to Jamie Dimon. *Medium* (2017). https://blog.chain.com/a-letter-to-jamie-dimon-de89d417cb80.

Jordan Pearson. 2016. When AI Goes Wrong, We Won't Be Able to Ask It Why. *Motheboard* (2016). https://motherboard.vice.com/en_us/article/vv7yd4/ai-deep-learning-ethics-right-to-explanation.

Taylor Pearson. 2017. The Blockchain Man. *Ribbonfarm* (2017). https://www.ribbonfarm.com/2017/10/10/the-blockchain-man/.

Ryan Selkis. 2018. Skin-in-the-Game Coins. *Medium* (2018). https://medium.com/@twobitidiot/skin-in-the-game-coins-da0afdfdc650.

Eden Shochat. 2017. This company will self-destruct after its ICO. *TechCrunch* (2017). https://techcrunch.com/2017/12/26/this-company-will-self-destruct-after-its-ico/.

James Somers. 2018. The Scientific Paper Is Obsolete. *The Atlantic* (2018). https://www.theatlantic.com/science/archive/2018/04/the-scientific-paper-is-obsolete/556676/.

Kai Stinchcombe. 2017. Ten years in, nobody has come up with a use for blockchain. *Hackernoon* (2017). https://hackernoon.com/ten-years-in-nobody-has-come-up-with-a-use-case-for-blockchain-ee98c180100.

Mustafa Suleyman. 2017. Exponential View, issue 137. *Exponential View, issue 137* (2017). http://www.exponentialview.co/.

Jason Vishnefske. 2016. IRIS A.I. Helps Us Make Healthy Chocolate. *Santa Barbara Chocolate corporate blog* (2016). https://www.santabarbarachocolate.com/blog/iris-ai-helps-us-make-heathy-chocolate/.

Gretchen Vogel. 2017. German researchers resign from Elsevier journals in push for nationwide open access. *Science* (2017). http://www.sciencemag.org/news/2017/10/german-researchers-resign-elsevier-journals-push-nationwide-open-access.

APPENDIX

Aiur Subsytems	PD 0	PD 1	PD 2	PD 3	PD 4	PD 5	PD 6	PD 7	PD 8	PD 9	PD 10	PD 11	PD 12
BLACKSTONE													
Hypothesis Extraction Engine													
MVP - problem identification			X										
M1 - identify hypothesis building information				X									
M2 - build hypothesis from text					X								
M3 - add 'black data'								X					
Knowledge Tree Builder													
MVP - identify single document building blocks						X							
M1 - identify recursive building blocks							X						
M2 - integrate hypothesis extraction										X			
Reproducibility Engine													
MVP - identify constraints and assumptions						X							
M1 - integrate black data								X					
M2 - auto test of experiments													X
Validity Engine													
MVP - check claims based on reproducibility									X				
M1 - text based flagging											X		
M2 - auto test of validity													X
ALEXANDRIA													
MVP - JAA + first version of the review platform					X								
M1 - integration of Blackstone										X			
M2 - add more journals												X	
CHAR													
MVP - keyword annotation tool		X											
M1 - hypothesis annotation tool				X									
M2 - balancing datasets								X					
PYLON													
MVP - reward mechanism for coders and QA			X										
M1 - self-deployment scripts core				X									
M2 - self -deployment scripts rest											X		
ALDARIS													
MVP - appeal service for reviewers and authors							X						
M1 - appeal service for coders and AQ									X				
M2 - general dispute resolution												X	
CBR-Governance													
MVP - minimum Constitution	X												
M1 - anti-fraud service	X												
M2 - consensus and decision making			X										
M3 - advanced anti-fraud											X		
CBR-Finance													
MVP - basic Institution	X												
M1 - PDF implementation + taxation service		X											
M2 - basic stability mechanisms		X											
M3 - Oracle + advanced stability mechanisms			X										

Fig. 13. Project Milestones vs Project Deliveries