Searchr - Pathing a route to Strong AI

Shaun Narayan

February 24, 2018

Abstract

The raw total volume of data doubles yearly. Such an overwhelming abundance of content has seen engagement rates drop, leaving pockets of content often undiscovered.

Searchr aims to connect relevant content with consumers, by qualifying search results using a novel one-click ML feedback loop built into the search interface.

Building useful Artificial Intelligence requires massive interdisciplinary collaboration. Our current forms of 'narrow AI' require underlying model features to be derived from raw data by domain experts. Network topology is often influenced by neuroscientific research into human physche, and performing online learning must be underpinned with a solid UX framework.

In the spirit of AI, ML, and optimization - our goal with Searchr is to methodically, and continually, refine and optimize our core operational processes towards the eventual goal of development of Strong AI (AKA Full AI, Artificial General Intelligence).

The following outlines our approach to do so. We have the great fortune to benefit from recent developments in the blockchain space which allow the development of an agile, highly adaptable and therefore optimizeable organisation (DAO).

- The Ethereum foundation [1]
- The developers of OpenZeppelin
- The Aragon team [2]
- The developers of Ganache/Truffle for a highly productive stack
- Giveth for their MiniMeToken standard [4]

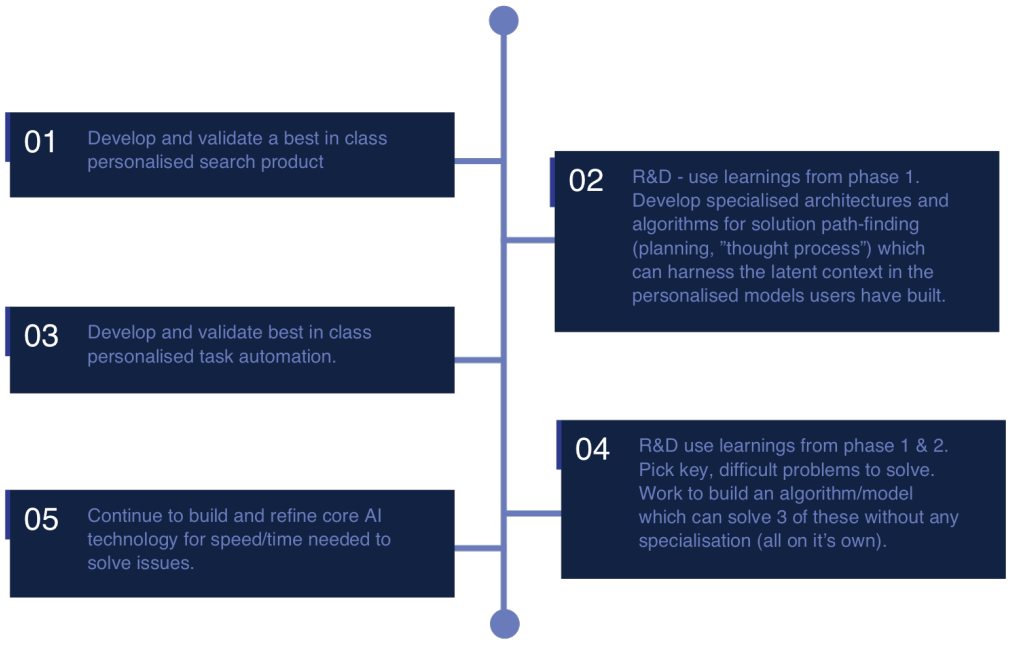
Contents

Chapter 1

Introduction

1.1 High level goals

The product roadmap will be developed over time to achieve Searchrs eventual targets, which are encompassed by these high level goals.



1. Develop and validate a **best in class** personalised search product
    - Form partnerships to increase the quality/relevance of search results (thematic, OpenAI, universities) - harness specialised knowledge, try to keep ourselves focused on the general aspect, there

are already key players who have developed solid models.

2. R&D - use learnings from phase 1. Develop specialised architectures and algorithms for solution path finding (planning, "thought process" simulation) which is some form of "time-aware" algorithm which can harness the latent context in the personalised models users have built.

3. Develop and validate best in class personalised task automation.

    - Feedback learnings from phase 2 into the community

4. R&D use learnings from phase 1 & 2. Pick key, difficult problems to solve. Work to build an algorithm/model which can solve 3 of these without any specialisation (all on it's own).

    - Crowdsource a list of key global, highly impactful issues we are facing
    - In tandem, use learnings to build strong safties, ethical guidelines etc.

5. Continue to build and refine core AI technology for speed/time needed to solve issues.

Chapter 2

The Searchr platform

Users interact with top level contract to create search jobs Jobs notify miners through the notification log Miners run the jobs, and commit results back to the blockchain (result URL + relevance ranking, with associated tags and resources)

Need to validate results to prevent spamming or fakes

When the validator contract verifies a result, it is added to the context store, and an appropriate portion of tokens are transferred to cover gas costs.

2.1 Rationale

Any such system requires massive data storage, bandwidth, and processing capacity. Currently, many large scale solutions face problems with economies of scale. Balancing the growth of supply and demand is a hard task, and often results in narrow AI by necessity of survival.

Using blockchain technology allows us to distribute this workload within an incentivisation scheme which supports sustainable growth.

2.2 Client

The Searchr client is React application built with truffle. Users can initiate searches to the service provisioning contract using metamask directly through the UI.

2.3 Smart contracts

The client interacts with the top level Searchr contract to issue a request for services to the miner pool with payment.

2.4 The SCHR Token

Consumers will be charged in SCHR (an ERC-20 token[3]) for their use of platform services. The service provisioning contract accepts payment, which will be forwarded to miners, with a portion reserved for long term product development.

- Map nodes (ETH GAS)[6] - 35%
- Pathfinder nodes - 60%
- Ongoing development - 5%

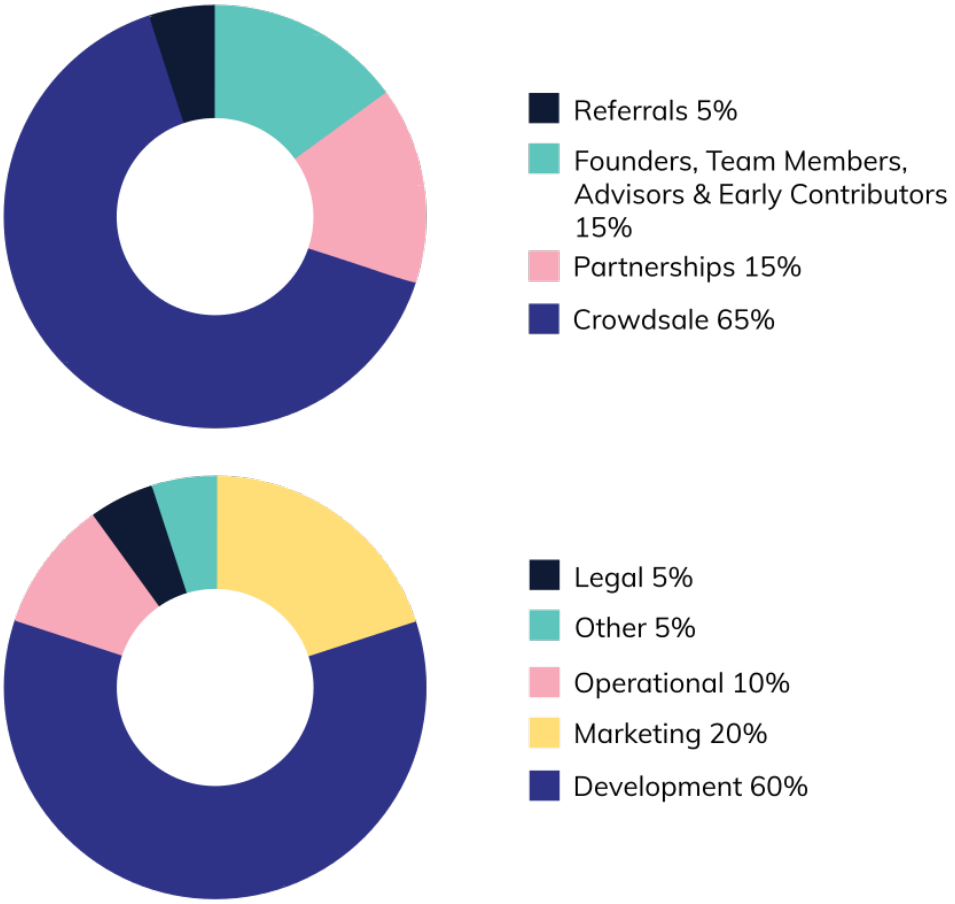
For services rendered.

2.4.1 Token Generation Event

All SCHR tokens will be minted at once, and no further tokens will ever be introduced[5]. The introduction of a burn rate is currently up for discussion.

- Total amount of SCHR tokens: 100,000,000
- Week 1-2: 100 SCHR = 1 ETH
- Thereafter: 66 SCHR = 1 ETH

Any unsold tokens will be burnt.



2.4.2 Token vesting

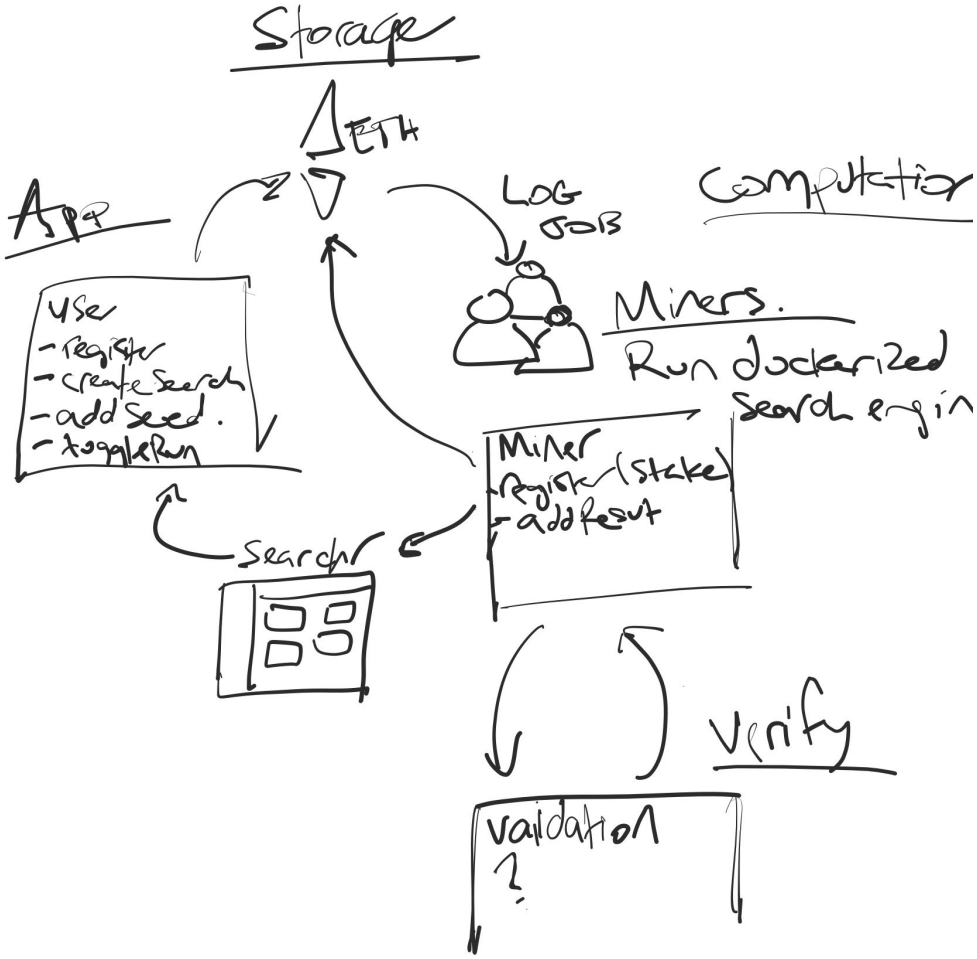
In accordance with best practise, all team members, founders, advisor and early contributors are subject to a 2 year vesting period with 6 months cliff to ensure long term commitment to the project.

2.4.3 MiniMeToken

SCHR is a MiniMeToken[4], which means that it is completely upgradable and able to be granted additional features as may be required over time, such as formalised governance voting on the aragon network.

2.5 Miners

Miners perform computation for searches, delivering results for requested researches in exchange for payment in SCHR tokens.



2.5.1 Reputation

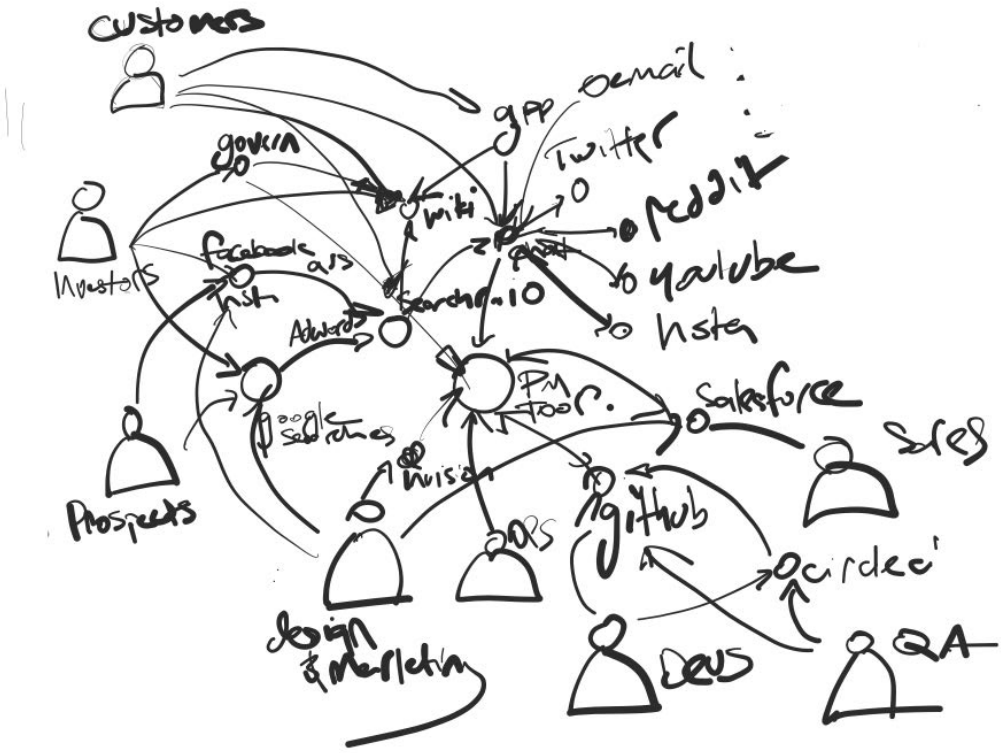
Every agent will be scored for reputation by network participants. Bad behaviour will decrement a reputation counter.

Chapter 3

Operations

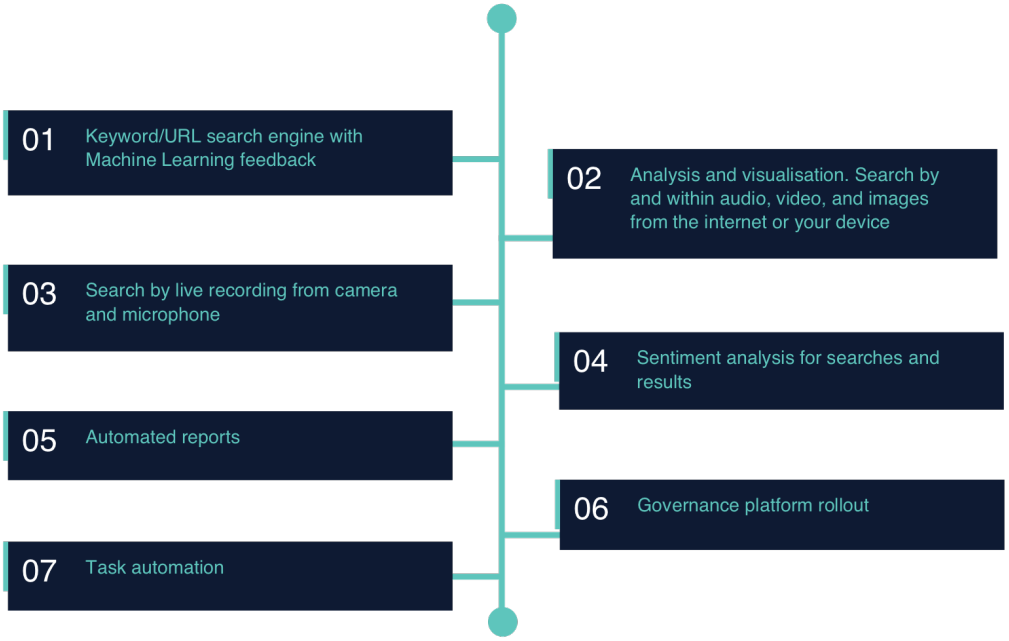
Searchr Ltd is a not for profit organisation registered in Estonia: [insert company number]. The organisation exists to provide ongoing platform development through hiring of developers for the purposes of delivering on our proposed roadmap.

As we operate as a remote team, an asynchronous workflow is incredibly important to reduce blockers. We use tools which facilitate agile, rapid development, and use integrations with Zapier connections to share information across tools.



3.1 Product roadmap

The current product roadmap looks to a 6-12 month horizon, with later stages being less concrete in scope and time estimates than those before.



3.2 Partnerships and Integrations

We don't want to reinvent anything. There are numerous platforms which we can leverage to accelerate development. As the team and product grows, we will look into opportunities for integration with some of the following platforms, where it is deemed appropriate following technical review.

Specialized providers

- Thematic
- OpenAI

Model marketplaces

- AIBlockchain
- SynapseAI
- Neuromation

3.3 Legal

We are committed to operating in a transparent and risk sensitive manner, taking all possible precautions to ensure the longevity and success of the project. We employ the services of LeapIN for financial support, including access to EU banking services.

We have also contracted the Argon Group to complete due diligence on our token offering.

This also necessitates care around anti money laundering requirements, and as such we perform KYC for all registrars.

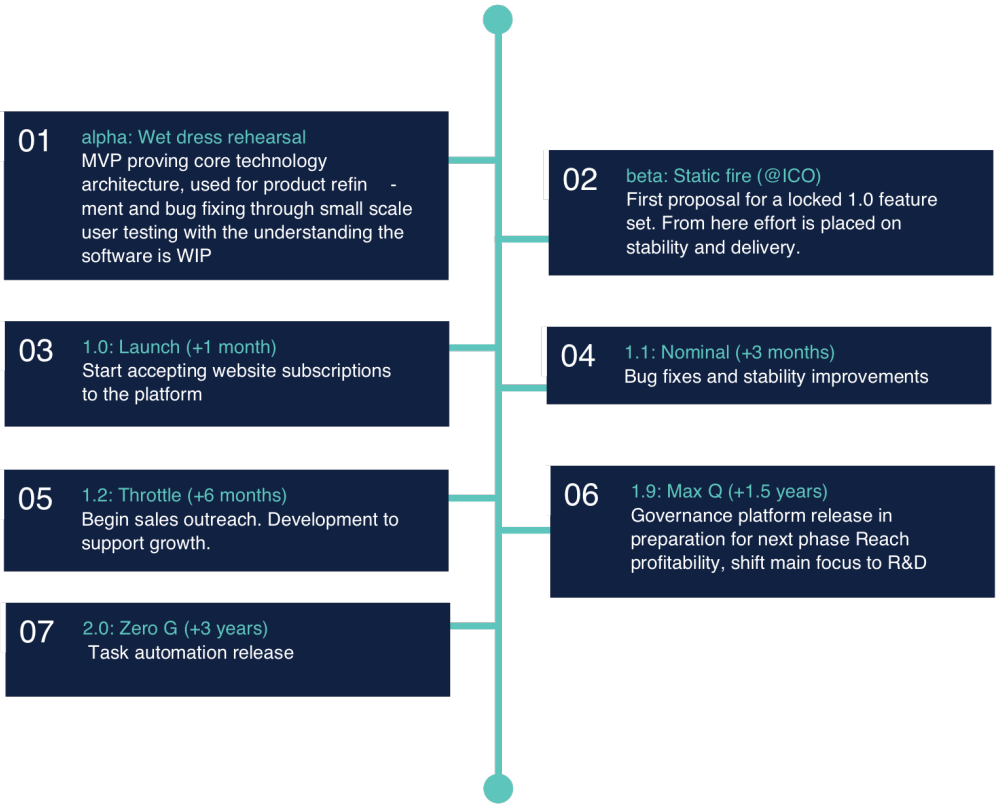
3.4 Risks

- Bank account shutdown
- Failure of the underlying network (Ethereum)
- Fall in token value
- Failing to curb bad actors

Chapter 4

Development

4.1 Development roadmap



4.2 Contributing

All source code is hosted on the Searchr repository. Pull requests are required to pass codeclimate tests before being considered ready for deployment to the release branch, which is then delivered to master weekly for deployment (where appropriate).

Our coding standards (linting), test coverage requirements and other such quality assurance measures are all automated through codeclimate.

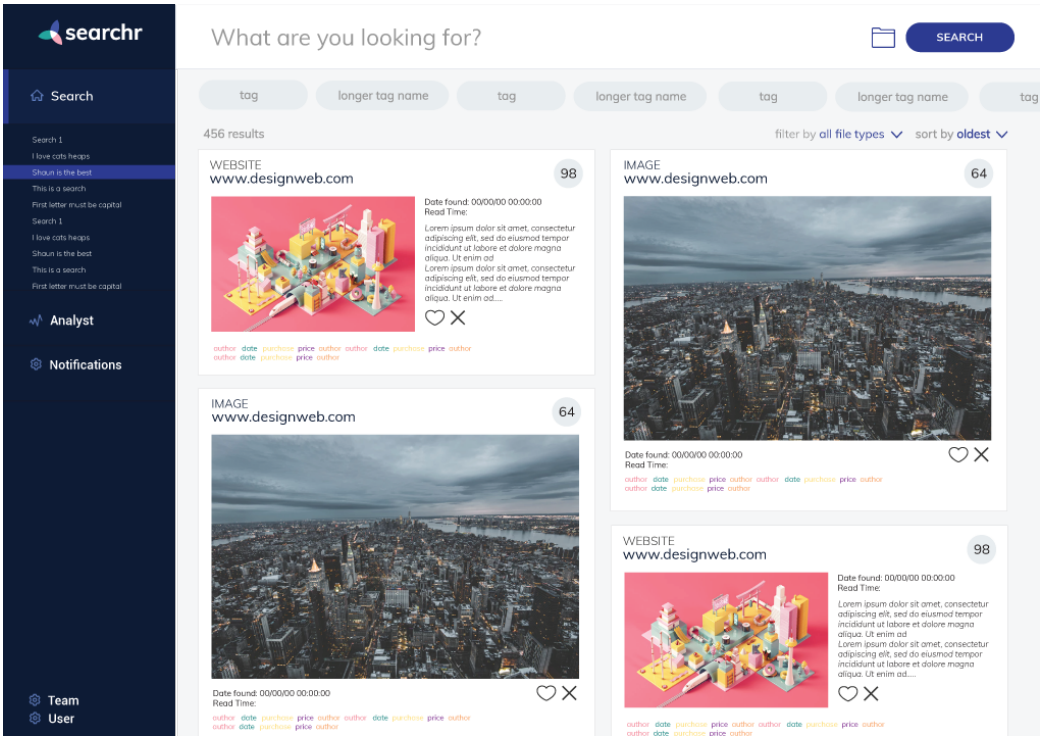
Chapter 5

User experience and design

We've already begun developing the look and feel for the client application. In the coming weeks, we will work closely with early adopters to refine the experience.

5.1 Search

The search screen has been designed to streamline the process of search refinement.



5.2 Analysis

We provide various visualisations of search results to assist in exploring complicated relationships among results.

[TBD]

Chapter 6

Strong AI

There's still a lot we don't understand about the human brain. At the same time, neuroscience is making huge strides, with advancements in medical imaging (FMRI, cell tagging), computational simulation power and more. We can now understand how the human visual pathways work with extraordinary detail (although, of course, there is still a lot not yet known).

We are slowly, but surely, reverse engineering our minds.

Searchr aims to advance research in the field of Strong AI in accordance with **consumer driven values**, by reinvesting and using acquired knowledge to further our understanding.

A trained autonomous agent is only as good as the training it was given, just as we are when we first learn to read. If the trainer imbues their personal bias in the underlying models, then the agent will behave accordingly (see Tay, Microsofts attempt at a crowdsourced intelligence chatbot). Therefore it is of upmost importance to build some form of reputation system to protect our common interests as humanity - a task suited perfectly for blockchain - for more information on reputation, see [insert section].

From a computational perspective, artificial neural networks, spiking neural networks and other existing solutions are shown to be functionally similar enough to our observations of human neural networks that they can be treated as effectively the same for the purposes of Searchr.

The process of shaping this computational "putty" of machine learning into something which responds to stimulus as a human would is known as feature mapping.

6.1 Feature mapping

Feature mapping is integral to many forms of AI prevalent today. Some of the most popular methods for extracting features from raw data:

- Manual/Hand crafting (adding domain specific knowledge)
- Reinforcement learning
- Particle swarm optimization
- Gradient decent (Autoencoding)
- Principle component analysis

As part of this process we are looking for a set of numbers, which when used as co-efficients within a ML model such as a neural network allows the network to signal the existence of a feature when presented as stimuli (an image, or text etc.), these numbers "tune" a network to actively respond in the presence of the mapped stimulus (a picture of your pet, for example).

This is somewhat similar to how our own neural networks organise in layers and map a series of simultaneous stimuli at one point in time to another series which is in turn provided as input to the next layer. This organisation of cells inspired deep convolutional neural networks.

6.2 Defining the problem to solve

Our problem statement is pretty simple, we want to implement strong AI, which can pass some form of "turing test" (TBD)

6.2.1 Defining a path to finding the solution

The current proposal for a solution approach is as follows

- **Context mapping**
  Structured data -¿ Searchr -¿ Structured results
- **Sensory mapping**
  Unstructured data -¿ Searchr -¿ Structured results
- **Task automation - muscle mapping**
  Unstructured data -¿ Searchr -¿ Actions

- **State -¿ action mapping, reinforcement learning** Time-variant, unstructured data ¡- Searchr (time aware) -¿ Actions (actuator responses - vocalisation, limb movement)

Structured data = JSON, XML, HTML
Unstructured data = Videos, audio, images, free text

By starting from the inside out, we can first define the problem in the context of existing solutions. First, we build algorithms which assist in forming a contextual encoding of knowledge - meeting feature parity with a human brain with the exception of

- sensory pathways/corticies
- muscle pathways/corticies

Sensory pathways will be short-circuited by injecting structured inputs (short term, specialised neural networks which transform and flow data between existing, pre-trained sensory models (Spacy, mxnet).

Muscle pathways will be intercepted and parse feature activations to structured outputs.

Layer to layer latency will not exceed 1000ms.

Next, we plan to build algorithms which can directly map (relatively) high resolution, high frequency, unstructured input signals to the previously developed context network. This will include decoding/encoding for bandwidth.

Layer to layer latency will not exceed 500ms.

Later, we will build inverse algorthims specialised for mapping context network activations to to actions (reinforcement learning inspired system).

Layer to layer latency will not exceed 100ms.

Finally, to complete the picture, a core planning algorithm will be built to consider inputs, context, and outputs over discrete periods of time in relation to the reward gained. Input will be transformed to a common N-D state space with time as the first dimension, and instead of feeding the system will employ actions to direct it's own sensors in the search space (becoming a self

acting agent).

Layer to layer latency will not exceed 20ms.

Chapter 7

Summary

[TBD]

7.1 Community/Support

If you'd like to reach us, we're ready to welcome you through any of the following channels;

- Rocket chat
- Reddit
- Twitter

Other useful resources

- Wiki
- Github

Bibliography

[1] V. Buterin *Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform.* https://github.com/ethereum/wiki/wiki/White-Paper

[2] L. Cuende, J. Izquierdo *Aragon network: a decentralized infrastructure for value exchange.* https://github.com/aragon/whitepaper/blob/master/Aragon%20Whitepaper.pdf

[3] F. Vogelsteller *"ERC: Token standard"*. https://github.com/ethereum/EIPs/issues/20

[4] J. Baylina *"MiniMeToken"*. https://github.com/Giveth/minime/blob/master/contracts/Mini

[5] J. Denne, C. Mair & J. Lambie *"Decentralized web services. A new era of cloud computing services, powered by blockchain technology."*. https://dadi.cloud/public/files/dadi.white-paper.pdf

[6] G. Wood *"Ethereum: A secure decentralised generalised transaction ledger"*. https://ethereum.github.io/yellowpaper/paper.pdf