AI's Role in R&D

Think Better.
Invent Faster.

Artificial intelligence has the capacity to vastly increase the output of research and development teams, and its introduction into the R&D process is an inevitability. Yet there have been no AI-based tools that have resonated with the R&D community. Brainstorming sessions take place on whiteboards that someone takes a picture of to email out to the group. Scanning the technical literature for inspiration or technologies to build off of, takes place through random, undocumented Google searches – if they occur at all. How is it that the people making the most innovative technologies are seemingly the slowest to adopt them in their workflow?

While AI is in the middle of a renaissance, software containing it has typically come in one of two forms, neither of which lend themselves to an R&D workflow

Black Boxes: Artificial intelligence algorithms can get so complicated, that a majority non-technical consumers do not need or even want to know what's going on underneath the hood. Thus most software with underlying AI does not spend any effort in explainability, and if anything plays up the "magic" of the process. R&D employees – technical and skeptical by nature – do not respond well to trusting that technology is working without any insight into it.

DIY Tools: For the more technical consumers, the supply of open source technologies such as TensorFlow or Keras gives the user significantly more control over and insight into the software's inner-workings. Yet this comes at the expense of time on the user's side, which makes its introduction as a workflow tool counterproductive.

In order for AI software to be effective in R&D, both its benefits and innerworkings must be made crystal clear to the user. While mystifying and idealizing AI is somewhat effective when marketing in certain fields, it has an extremely negative effect on engineers, researchers and scientists who regularly interact with cutting edge technology.

In this white paper, we lay out two very clear use cases and associated benefits of artificial intelligence in the research and development process.

Categorize Technology

Technical text – everything from small snippets to dissertations – must be ordered or classified in some way in order to get the most value from it. Similar technologies can be quickly scanned during the initial research process, work within an organization doesn't get duplicated across different teams/offices and technology evolves to its maximum potential.

Unfortunately, the dominant form of technical text classification is self-tagging – e.g. a patent attorney using a CPC code to classify a patent to the correct examiner or a graduate student making sure their paper gets in the right conference track. This self-classification causes a number of glaring issues

- Manual classification *always* results in errors

- There is virtually no way to connect two different systems of classification

- Even within a system, classifications get added, removed and changed over time

Conversely, AI categorization algorithms are **automatic**, **universal** and **dynamic**. In other words:

- Text is **automatically** classified based on it's content, not an exogenously applied tag

- Classifications are **universally** applied to text chunks between corpuses, and

- The precise classifications are **dynamic** in nature, so that a corpus can always be reclassified when under a given system

These characteristics of an AI classification system allow for two main benefits to R&D teams and managers: the clustering of similar technologies and decomposition of text into its composite technologies.

Clustering Similar Technology

As briefly mentioned, static classification of technical texts results in such classifications to become obsolete very quickly. For example, linguistics papers from 30 years ago would have never been classified under "artificial intelligence". Yet today, natural language processing is one of the largest subsets of AI and many linguistic technologies are vastly connected to new prediction algorithms. This type of common mis-labelings has resulted in classification systems rarely being used for anything other than administrative purposes.

Classification – both labeled and unlabeled – is the result of many AI algorithms. Without relying on previous technical classification systems, data points can be classified in the same bucket based on their content. For example – as one's conception of "artificial intelligence" starts to include syntax analysis – so too would the bucket of AI classified texts expand to linguistics papers.

Legit allows our user base to define their own classification systems based on their industry. Hardware companies define general categories such as "Automatic Doors" or "Batteries", while R&D teams working on medical devices can get as granular as "Epicardial Locomotion" and "Raman Spectroscopy". Both the classification of text and the definition of the classification itself are dynamic, leading to a highly useful system with which R&D teams can easily explore their idea landscape to maximize the value of their technologies.

Decomposing Ideas into Composite Technologies

Any new product or idea is composed of many different underlying technologies. Take the following description:

> an internet based application that uses predictive analytics to uses
> sensors to detect how a patient with a neurodegenerative disorder
> is mentally and physically deteriorating

There are at least three different technologies latent within this one product:

- Neurodegenerative disorder

- Predictive analytics

- Patient sensing and detection

Researching all of this one product's composite technologies takes hours of an R&D employees time typing out individual 3-4 word queries in Google scholar and other search engines. Furthermore, there are likely other technologies – such as internet-based medical applications – that are missed by the engineer.

In the same way that AI-based classification systems group together similar technologies, they can help to piece apart all different technical aspects of piece of text. Using the example above, an engineer who is attempting to build a system for detecting neurodegeneration could investigate other similar products in predictive analytics or discover that they have stumbled upon an innovative new way to sense a patient's physical movement.

This class of algorithms is what Legit uses to help their users deeply explore their ideas. A user can take the neurodegeneration query above and increase or decrease the presence of certain technologies in the results. Say for example that the an innovation engineer is particularly interested in other uses of predictive analytics in medical devices. Rather than change their query over and over, they can simply press the increase button next to "predictive analytics" to see technical chunks of text that are more in the direction of predictive analytics in medical devices. Not only does this save time, but it gives the engineer much more control over their exploration – as well as documenting their thought process for R&D managers and any future employees that need to inherit the project later on.

Using AI to group together similar technologies and separate individual chunks of text into its composite parts is one step further in creating a dynamic system in which the available data provides the maximum value to the use with the minimal extra workload.

Increasing Usability

Applications of AI are often times overly focused on generating insight from datasets, but not enough attention is given to the technology's ability to vastly improve the user experience (UX). Particularly in R&D – where users are inherently skeptical of a set of algorithms' ability to perform cognitive tasks such as insight generation – notable UX improvements can result in high dividends. To an R&D employee, showing an increase in software usability is significantly easier than proving the accuracy rate of a matching algorithm.

There are two main avenues through which artificial intelligence can improve UX for R&D teams, primarily in a exploratory capacity: deeper understanding of user input and better presentation of information.

Deeply Understanding User Input

Engineers, researchers and scientists often find themselves writing about the same technology in multiple different contexts. This results in a lack of documentation and investigation because the process of getting one's idea down on paper for recording process is completely separate from the exploration of similar ideas – either internal or external to the company. In other words, if I need to a) document my thought process b) inform the higher-ups what I'm working on and c) find technologies similar to the one I'm working on, then something at least one of those is going to suffer

Understanding different aspects of technical texts is an area in which AI – particularly natural language processing – excels. An R&D employee can submit a 2-3 long sentence that simultaneously:

1. Runs a number of queries based on its composite technologies

2. Extracts high-level topics to broadcast around the company what they're working on

3. Documents their thought process for retrospective analysis

For Legit, we accomplish this through semantic decomposition of user text – which allows us to understand many nuances of user's input. Rather than submitting multiple 3-4 word queries, half-heartedly tagging documents with categories and emailing scattered notes across the company, we perform all of the tasks immediately using artificial intelligence.

Intelligible Presentation of Technical Information

Reviewing prior technical ideas is often times overlooked by many R&D employees, particularly those closer to the engineering fields. This is a huge missed opportunity, as those in innovation engineering could greatly benefit from looking at past technologies – even those found outside of the engineering field.

Reading through technical documents, especially academic articles and patents, it extremely time consuming and often times daunting. While some R&D employees have been through graduate school and have experience with that type of work, many in the R&D field either do not like or are not used to reviewing long documents, where a majority of the sections are irrelevant to their work.

Aside from bringing back more relevant documents using semantic matching, AI and NLP algorithms can aid the user by making the technical information much more legible. Legit does this in two primary ways:

1 We present the *exact* part of the long technical document that we deem to be relevant to the user, thus sparing them the process of searching through a potentially long and irrelevant document for relevant information. This saves each R&D team member an estimated 474 hours annually!

2 By repurposing parts various parts of a technical document, we can re-describe particularly convoluted parts of a document using more intelligible language. This is particularly useful when looking at independent claims in patent documents.

By improving the user's experience of documenting their ideas and exploring similar technologies, the amount of data being aggregated in one place vastly increases. In turn, this vastly increases the amount of possibilities for artificial intelligence.

Conclusion

AI's introduction into the research and development process is inevitable, but its initial applications must be both transparent and easy to use. Otherwise, tech savvy R&D teams will either not trust the product's efficacy or find it burdensome to use it on a regular basis.

While there are many potential benefits of artificial intelligence in R&D, the two best examples are ones that directly confront the pain points of users. Technology is notoriously difficult to classify, so using AI to dynamically sort and categorize text is of immediate use. Increasing usability of R&D tools is of obvious use, as engineers, researchers and scientists generally look at past technologies less than they should because the process is so cumbersome.

Artificial intelligence has the capacity to revolutionize virtually every industry, R&D is no different. Through thoughtful introduction, modern companies can harness the new technologies to supercharge their innovation processes and fully leverage all the benefits that AI has to offer.

Legit is a Cambridge-based company that makes AI-powered software for R&D teams, managers and C-suite level executives. More information available at legit.ai