

Amplify: self-evolving human-in-the-loop



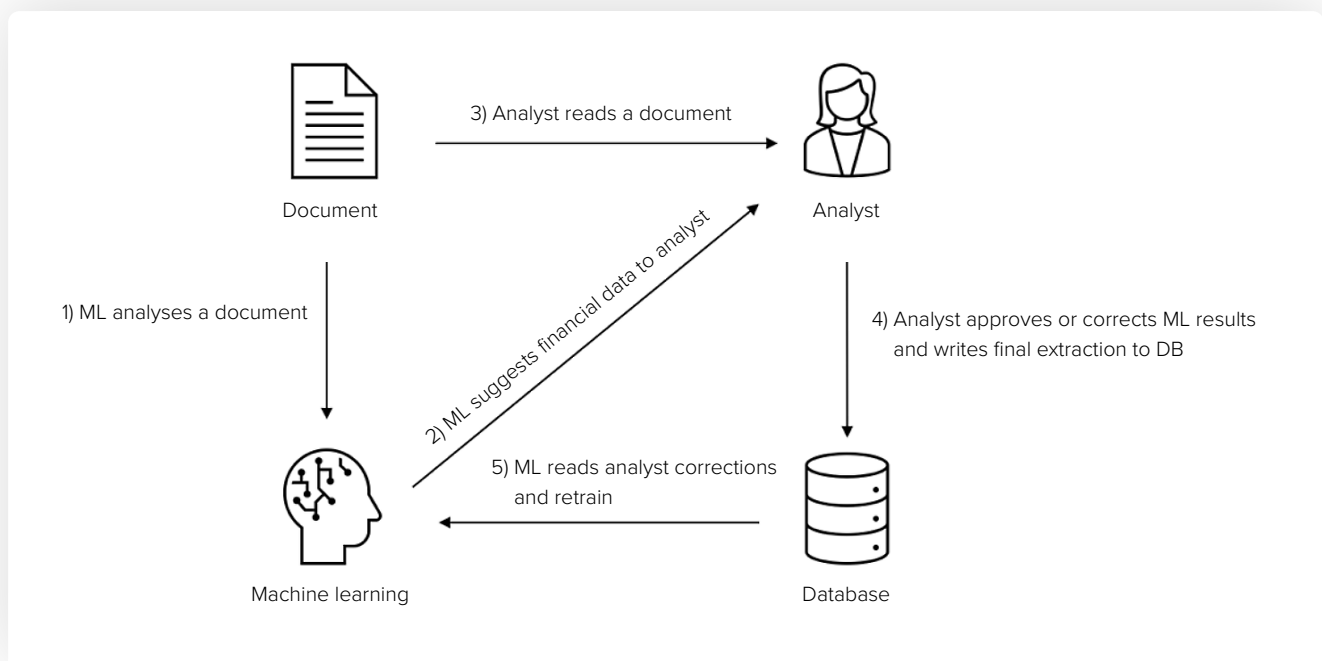


Introduction

A large part of what makes LSEG's offerings so valuable is its army of analysts sifting through countless documents and extracting financial data. This includes financial metrics from filings reports and ESG metrics from company reports – to name just two. The work is by and large manual and complex but LSEG have developed and deployed a human-in-the-loop (HITL) system to assist the analysts by partially automating their work.

Amplify, the HITL system, uses Machine Learning (ML) to train and refine models incorporating the analyst's expertise as a natural part of their workflow. The trained models automatically suggest to the analyst the financial data to be extracted in a specific document. This way, The analysts have an ever-self-improving system that helps them in automating their day-to-day job.

Figure 1: Amplify conceptual diagram

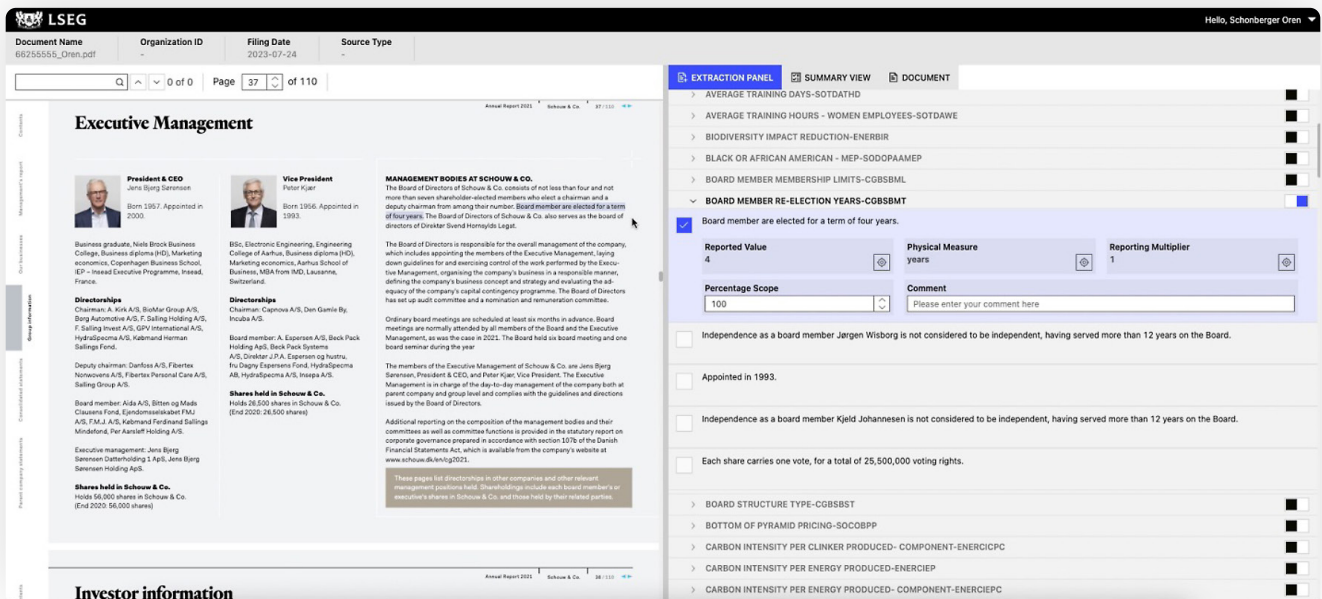


HITL in action

When analysts scrape ESG-related data from company reports, one of the metrics of interest is Board Member Re-election Years, i.e., the length of a board member’s term. To find the relevant figure, an analyst would open the document and look for key words such as ‘board’, ‘elected’, ‘term’, etc. Then they would go through the sentences containing the relevant key terms to determine if there is in fact a reference to the length of the board member’s term.

Amplify acts as an add-on to this flow, where instead of searching for key terms, an analyst can view a list of suggested ESG metrics by the system (right-hand side in screenshot below). Once she presses on an item in the list it opens and highlights to the left the relevant sentence from the document as well as showing to the right the actual values:

Figure 2: Amplify - an ESG overview



Source: LSEG, Nov 2023

The analyst may choose to accept, amend or reject the system’s suggestion. This choice will contribute to improving the model and lead to a better suggestion in the next document.

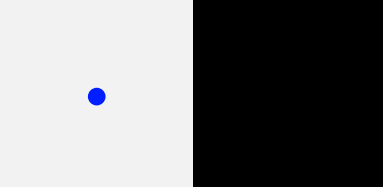
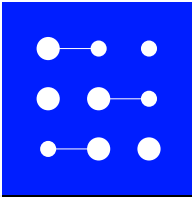
ML 101

One of the benefits of Amplify is its ability to improve upon the analyst’s feedback (approve, amend, reject). This is done using ML and Large Language Models (LLMs).

A very basic notion of ML is discriminating between two classes, e.g., Board Members Re-election Years vs. all the rest. To learn to do this, during ‘train time’ an ML model must be presented with examples representing each of the two classes. These examples are obtained in Amplify free of charge, since we use analysts’ work to collect them implicitly – i.e., positive examples are taken from those sentences analysts have flagged as related to the class of Board Members Re-election Years while negative ones are taken from the rest of the sentences.

Once we have the set of examples, we need to encode them in a way that will let the ML model train easily. There are lots of ways for doing that, e.g.: character-based encoding, word-based encoding, custom rules – sentence contains one of a predefined list of words, sentence contains numbers.

One of the challenges with these methods is that they need a reasonably large set of examples to kickstart the model – but for the system to be effective from day one, we want it to work on a few-shot example set.



The value of LLMs

LLMs are trained on vast amounts of text and are good at encoding sentences into numeric vectors in a way that allows an ML model to discriminate between classes using as little as a single example.

So, the full flow is as follows:

1. At train time, all the examples are encoded using an LLM; then an ML model is trained to discriminate between the two classes.
2. At runtime, all the sentences of the current document are encoded using the same LLM and fed to the model trained in stage (1) which identifies each sentence as Board Members Re-election Years-related or other.

Technicalities under the hood

One advantage Amplify has over other ML-based projects is the fact that training happens iteratively and automatically in the production environment. This calls for a significant [MLOps](#) consideration.

In Amplify we use a cloud-based approach to facilitate the following:

- Iterative and automatic model training, validation and deployment
- GPU provisioning: the usage of LLMs requires specialised hardware (GPUs) to ensure acceptable latency

Some questions to consider

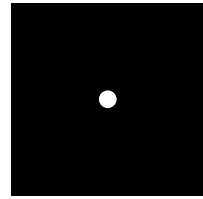
When building Amplify, we considered the following questions:

- Model sharing: there are numerous analysts performing the same work on different documents. Should they all train and use the same model or should they share a joint one? By sharing a model they can all enjoy the benefits of each other's work... but one rogue analyst can compromise the effectiveness of the rest.
- When to trigger model training? In theory, it makes sense to train a model every time new analyst feedback is available. But model training is costly in terms of compute resources and time, so we need to consider the ROI of each training step.
- How do we make sure that the new model is better than the previous one? In theory more data means a better model – but in practice some data may be problematic and cause the model to misbehave.
- What should be done with concept or data drifts? Over time, the documents may change their nature as well as some concept definitions. When such a drift occurs, we need to identify it and decide whether we want to mix data from before the drift with new data.

Conclusion

At LSEG, we place great emphasis on collecting and supplying our customers with high-quality data. This information gathered by thousands of experienced analysts who feed our databases with industry-leading data. Amplify was created using our extensive experience in NLP and ML to make their work easier, allow them to concentrate on the more difficult tasks and expand our coverage. It is a solution designed to empower both our analysts and our customers.





The content of this publication is provided by London Stock Exchange Group plc, its applicable group undertakings and/or its affiliates or licensors (the "LSE Group" or "We") exclusively. Neither We nor our affiliates guarantee the accuracy of or endorse the views or opinions given by any third party content provider, advertiser, sponsor or other user. We may link to, reference, or promote websites, applications and/or services from third parties. You agree that We are not responsible for, and do not control such non-LSE Group websites, applications or services.

The content of this publication is for informational purposes only. All information and data contained in this publication is obtained by LSE Group from sources believed by it to be accurate and reliable. Because of the possibility of human and mechanical error as well as other factors, however, such information and data are provided "as is" without warranty of any kind. You understand and agree that this publication does not, and does not seek to, constitute advice of any nature. You may not rely upon the content of this document under any circumstances and should seek your own independent legal, tax or investment advice or opinion regarding the suitability, value or profitability of any particular security, portfolio or investment strategy. Neither We nor our affiliates shall be liable for any errors, inaccuracies or delays in the publication or any other content, or for any actions taken by you in reliance thereon. You expressly agree that your use of the publication and its content is at your sole risk.

To the fullest extent permitted by applicable law, LSE Group, expressly disclaims any representation or warranties, express or implied, including, without limitation, any representations or warranties of performance, merchantability, fitness for a particular purpose, accuracy, completeness, reliability and non-infringement. LSE Group, its subsidiaries, its affiliates and their respective shareholders, directors, officers employees, agents, advertisers, content providers and licensors (collectively referred to as the "LSE Group Parties") disclaim all responsibility for any loss, liability or damage of any kind resulting from or related to access, use or the unavailability of the publication (or any part of it); and none of the LSE Group Parties will be liable (jointly or severally) to you for any direct, indirect, consequential, special, incidental, punitive or exemplary damages, howsoever arising, even if any member of the LSE Group Parties are advised in advance of the possibility of such damages or could have foreseen any such damages arising or resulting from the use of, or inability to use, the information contained in the publication. For the avoidance of doubt, the LSE Group Parties shall have no liability for any losses, claims, demands, actions, proceedings, damages, costs or expenses arising out of, or in any way connected with, the information contained in this document.

LSE Group is the owner of various intellectual property rights ("IPR"), including but not limited to, numerous trademarks that are used to identify, advertise, and promote LSE Group products, services and activities. Nothing contained herein should be construed as granting any licence or right to use any of the trademarks or any other LSE Group IPR for any purpose whatsoever without the written permission or applicable licence terms.

Copyright © 2024 London Stock Exchange Group. All rights reserved.

Visit lseg.com |  @LSEGplc  LSEG



LSEG DATA & ANALYTICS