

PAPUD

A unified approach to heterogenous data

Large Language Models have become a ubiquitous presence in our daily lives since their public debut, notably with the introduction of ChatGPT in November 2022. However, the attempt to extract meaningful insights from diverse datasets is not new; it has been an innovation topic for many years already. The ITEA project PAPUD exemplifies this pursuit. Running from 2018 to 2020, PAPUD united 16 partners from Belgium, France, Romania, Spain and Türkiye with the aim to empower companies in exploiting their large amounts of heterogenous data using Deep Learning.

Businesses are faced with a huge variety of autonomous, heterogeneous data sources - from social media to Internet of Things. The corresponding 'data deluge' is too much for most to handle, yet almost every industry can benefit from the competitive insights that Deep Learning-based data analysis can unlock. Recognising that the value of Deep Learning lies not in independent analytics processes but rather in a unified approach to different types of heterogenous data, PAPUD

(Profiling and Analysis Platform Using Deep learning) therefore created a unique software platform and new Deep Learning algorithms to optimise the processing of this data. Five use-cases demonstrated the project's success: e-Commerce, Call centre operations, Recommendation system for human resources, Behaviour analysis for reverse efficient Modelling, and Prescriptive maintenance for High Performance Computing (HPC).



PAPUD's technological innovations were divided between the hardware of the platform, the Deep Learning software and the domain-specific use-case tools. The process began with the acquisition of data from diverse sources such as surveys, reviews or calls; this was fed into the platform via application programming interfaces and adaptors. Data could then be defined and characterised through HPC infrastructure and AI tools like TensorFlow, PyTorch and other open-source software libraries. In order to provide a complete application, PAPUD took privacy into account, with the use of Docker to bundle data into separate containers that allow partners the exclusive protection and control of their own information.

Following the pre-processing, the Atos-hosted PAPUD platform carried out the data analysis. A combination of Deep Learning, Machine Learning and Data Mining tools, libraries and resources produced models which were stored and visualised for end-users via a dashboard. The result was a series of recommendations which businesses

could use to optimise or improve their processes and services.

A concrete example was the Call centre operations use-case, for which KU Leuven had developed Deep Learning text models. 4C Consulting integrated these into its AI platform TellMi which automates the analyses of all text-based customer interactions across different languages, in real time. 4C's objective was to help companies become customer centric and, by extracting the insights from text-based customer interactions, understand their customers better as well as improve service provision. TellMi has been offered as a consulting service and a standalone product:

1. **Consulting mode:** 4C offered consulting services to train and apply models to provide an overview of the deep insights hidden in text. This is often a one-off analysis and an easy way for customers to dip their toes into the wonderful world of AI.
2. **Product mode:** The TellMi platform was offered as a self-service AI product which customers can easily use to train models to extract deep

insights across different languages themselves. TellMi can be fully integrated into the customer's work environment.

In July 2020, 4C was acquired by Wipro, a leading global company in the field of information technology, consulting and business process services. The acquisition significantly strengthened Wipro's position as a leading provider of salesforce solutions in the United Kingdom, France, Benelux, the Nordic countries and the United Arab Emirates regions, where 4C already had a strong position.

In tangible terms, PAPUD's main contribution to businesses is greater efficiency achieved through sizeable improvements in Deep Learning. For example, the Area Under the Curve (AUC) – the ability of a classifier to distinguish between classes – stood at 0% for keyword extraction at the start of the project in 2018 but is now 93.7%. Similarly, the accuracy of Deep Learning-based models for HPC prescriptive maintenance was increased from 50% to 95%.

Project start
January 2018

Project end
December 2020

Project leader
Cornel Crisan
Bull, France

More information
<https://itea4.org/project/papud.html>

For HI Iberia Ingenieria y Proyectos, these kinds of improvements have cut the time taken to find a perfect match through CV processing from five days to three. The tool developed in PAPUD continues to be actively utilised by HI-Iberia's HR department. Through recent updates integrating the latest advancements in Deep Learning for Natural Language Processing, this tool plays a pivotal role in identifying fresh talent within highly competitive domains such as ICT. Its continued evolution enables the company's department to stay ahead, effectively scouting and nurturing talent in this rapidly evolving sector.

Other beneficiaries of PAPUD include Turkgen, which is very active in the fintech domain with its virtual assistant solution CBOT. Over the past decade, this virtual assistant solution has been implemented in huge Turkish banks with more than 10 million customers. A huge amount of data is coming to

those virtual assistants; processing this quickly and correctly is becoming more critical than ever. In the PAPUD project, Turkgen had the chance to collaborate with its partners' Turkish text data and this helped Turkgen to improve its Turkish text mining algorithm, which is part of Turkgen's current main solution. Over time, the business has improved and the team that is working on it has increased in parallel. Currently, the biggest banks of Türkiye and many big companies and organisations prefer Turkgen's virtual assistant solution for their customers, including İşbank, Ziraat Bank, MediaMarkt, McDonald's, Türk Telekom, PepsiCo, Bayer and İstanbul Metropolitan Municipality. The PAPUD project therefore helped Turkgen a lot.

Pertimm developed an AI-based recommendation module that serves as an added value module for its existing e-commerce platform Pertimm Search suite and which takes into account data from baskets. Currently, this module has been sold to several customers but serves especially as an incentive to sell the Pertimm Search suite which is one of the company's main businesses. A sales boost has allowed Pertimm to hire two new engineers.

In addition, Atos has integrated some of the results into its Codex AI Suite to tackle the most resource and performance demanding use-cases. One aspect is overheating: with PAPUD, 70% of overheating events can be predicted and preventive actions can reduce the costs of this by 65%.

The results of PAPUD also provide precious input for Eviden, an Atos Group company, as it has been integrated into the company's Proactive Maintenance product, which is part of Eviden's Smart Maintenance Management Suite. The targeted use cases addressed by this product will contribute to predicting and anticipating key and complex issues happening in the HPC world, such as interconnect contention, overheating, and energy savings. The value brought to the market by this product will be available by November 2024.

Less tangibly, PAPUD also represented a message to industry as a whole:

PAPUD made it significantly easier for companies to benefit from data analytics. Use of the PAPUD platform and the creation of domain-specific Deep Learning tools allow businesses to bypass huge organisations which have dominated the field, saving them time and money while also allowing them to tailor resources more specifically to their own internal or commercial needs.

PAPUD's activities were primarily centred around predictive AI, a prevailing subject during the project's lifetime. However, several of PAPUD's findings are now transferable to the latest wave of generative AI. Firstly, the PAPUD platform, comprising both software and hardware, utilises containers to encapsulate computing tasks and GPU for enhanced performance, making it highly suitable for any generative AI application requiring a controlled execution environment and intensive processing capabilities. Secondly, the expertise acquired in customising general models to specific objectives is very relevant for generative AI, where the business will be created by exploitation of these large models for specific tasks.

In conclusion, the PAPUD project yielded direct benefits for its partners in the predictive AI domain and has well prepared them to leverage their know-how in the context of the new generative AI applications wave. Ultimately, PAPUD has demonstrated that greater efficiency translates into cost savings and increased sales, and so provides an opportunity for businesses of all sizes.