IDP: Developing an AI-Software for automated Finance Reporting and Data Analytics at Parabella

03.06.2024

Project: Development of an AI-software that extracts finance documents and uses the financial numbers to derive analytical insights and create a finance report automatically with a “one click to report” feature. These automated finance reports will be multi-validated in an AI-pipeline to become audit-aligned with the institutional frameworks.

1. Motivation and Background

Transparent communication of annual financial numbers is essential for every company, investor, regulator and stakeholders. The data input for a traditional finance report is based on manually data collection and often not automated, especially the textual interpretation of the numbers including forecasts, risk assessments and so on. Parabella Analytics aims to automate this whole reporting process by applying state-of-the-art AI techniques by extracting necessary data automatically, represent analytical insights and converts all financial data into coherent, narrative text of a finance report.

2. Project Objectives

- **Algorithm Development:** Identification of a data-to-text algorithm like BART or T5 to build a foundation framework that is specified on financial reporting. Using training data to test efficiency and outcome-quality.

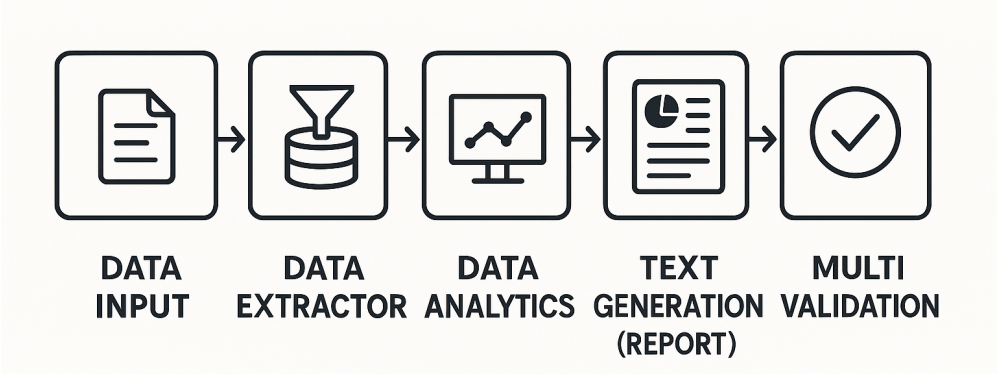
- **LLM Integration & Fine-Tuning:** Evaluation of an AI-Architecture between BART algorithm and GenAI models. Evaluate and text large-language models to translate numeric inputs into coherent text at paragraph and section levels.

- **Data Analytics & Schema Design:** Define, how extracted, numeric data can be used to build analytics frameworks and graphics. Identify a dynamic report schema for multiple topics e.g. Management Discussion & Analysis, KPI commentary, risk management, heatmap and so on. Creativity is important.

- **End-to-End Automation:** Design and implement a “one-click to report” workflow that ingests raw financial statements and analytics-outputs in form of a structured narrative report, without human intervention. You can use the current AI-pipeline of the ESG-reporting module from Parabella Analytics.

- **Data Validation & Consistency Checks:** Incorporate rule-based and AI-driven multi-validations with different GenAIs to ensure factual accuracy, flag anomalies and maintain regulatory compliance for „audit-aligned reports“

  AI pipeline methodology:



Framework: AI pipeline for automated finance reporting and data analytics for audit-aligned report

Interdisciplinarity: The project merges core Informatics competencies (software development, data management, AI/ML) with expertise from the application subject (e.g., Strategy and Organization, Business Informatics, Finance & Accounting, Sustainability Management). Students will apply informatics methods to a real-world business problem in the rapidly growing ESG sector, considering strategic implications for companies.

3. Planned Methodology (Scrum) and Technology Stack (Proposal)

The project will follow the Scrum agile development framework to ensure iterative progress, flexibility, and continuous feedback. The project duration is 3 months (full-time).

- **Scrum Roles:**
  - **Product Owner:** Alexander Pavlovski and Marcus Ohnemus. Responsible for defining the product vision, managing and prioritizing the Product Backlog.
  - **Scrum Master:** One of the students (can rotate) or the academic supervisor if they are actively involved. Facilitates the Scrum process, removes impediments.
  - **Development Team:** The 3-4 students. Self-organizing team responsible for delivering a potentially shippable increment of the product at the end of each Sprint.
- **Scrum Events:**
  - **Sprints:** Fixed-length iterations of **2 weeks**. (For a 12-week project, this means 6 Sprints).
  - **Sprint Planning:** At the beginning of each Sprint to define the Sprint Goal and select items from the Product Backlog for the Sprint Backlog.
  - **Daily Scrum:** Short daily meeting to synchronize activities and create a plan for the next 24 hours.
  - **Sprint Review:** At the end of each Sprint to inspect the increment and adapt the Product Backlog if needed. Stakeholders (Parabella Analytics) will participate.
  - **Sprint Retrospective:** After the Sprint Review and prior to the next Sprint Planning, to reflect on the past Sprint and identify improvements for the next.
- **Scrum Artifacts:**
  - **Product Backlog:** An ordered list of everything known to be needed in the product. Managed by the Product Owner.
  - **Sprint Backlog:** The set of Product Backlog items selected for the Sprint, plus a plan for delivering the product Increment and realizing the Sprint Goal.

    - **Increment:** The sum of all Product Backlog items completed during a Sprint and all previous Sprints. A "Done" increment is usable and potentially releasable.
- **Definition of Done (DoD):** A shared understanding of what it means for work to be complete, ensuring quality. To be defined by the team at the start.
- **Technology Stack (Proposal):**
    - **Backend:** Python (with frameworks like Django/Flask).
    - **Database:** PostgreSQL or similar.
    - **Frontend:** Web technologies (e.g., React, Vue.js, or Streamlit for rapid prototyping).
    - **AI/ML:** Python libraries (scikit-learn, TensorFlow/PyTorch, spaCy/NLTK).
    - **Version Control:** Git.
    - **Project Management:** JIRA, Trello, or similar for managing backlogs and sprints.

4. Expected Results and Deliverables

- A functional software prototype of an AI-pipeline for automated finance reporting and data analytics
- A unique AI-architecture with a self-developed and finetuned BART algorithm
- A well-maintained technical documentation of the architecture, algorithm, GenAI interaction, data models and API descriptions
- A well-maintained product backlog and sprint backlogs
- Regularly demonstrated sprint reviews
- A detailed final project report outlining the process, Sprints, results, and limitations
- A final presentation/demonstration of the project and its outcomes.
- Source code of the developed prototype.

5. Task Distribution within the Team (Self-Organizing Scrum Team)

While students may have individual strengths (e.g., one focuses more on AI/ML, another on backend, another on frontend/UI, and one on ESG domain integration & testing), the Development team is cross-functional and self-organizing. They will collectively decide how to best accomplish the work in each Sprint Backlog. All team members are jointly responsible for project planning, documentation, and

delivering the sprint goals. The timeline is based on 6 x 2-Week Sprints over 3 Months.

6. Supervision and Resources

- **Product Ownership & Industry Supervision:** Alexander Pavlovski regarding AI and tech expertise, Marcus Ohnemus regarding finance input and business development.
- **Academic Supervision:** Prof. Isabell Welpe
- **Resources from Parabella Analytics:** Provision of domain expertise, software access to relevant tools and literature. Regular availability for sprint reviews and backlog refinement.
