Universiteit van Amsterdam

Using Generative AI in the Legal Domain

Group D1

13-12-2024

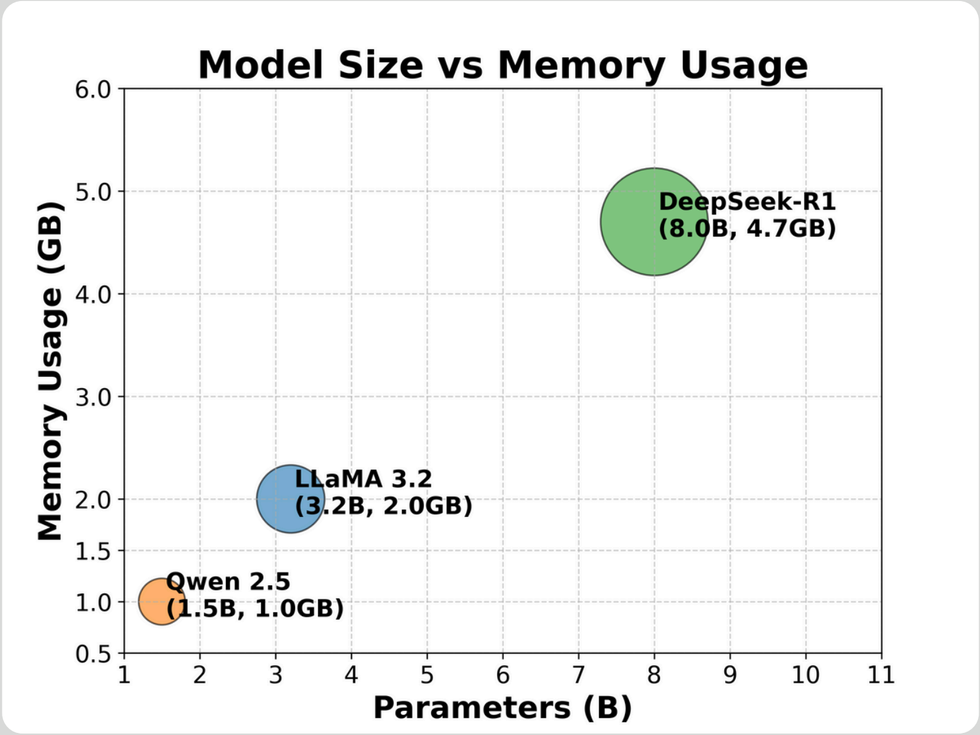
Team Di

Hybrid Decision-Making System for Dutch Student Finance

- **PROJECT CONCEPT: COMBINE LARGE LANGUAGE MODELS (LLMS) WITH DECISION TREES TO INTERPRET AND APPLY STUDENT FINANCE REGULATIONS.**
- **RELEVANCE: SIMPLIFIES COMPLEX REGULATIONS AND ENSURES PRACTICAL USABILITY FOR USERS.**
- **MODEL'S TRANSPARENCY: USE DECISION TREES TO PROVIDE CLEAR REASONING PATHS AND VISUAL EXPLANATIONS FOR DECISIONS.**
- **CONTEXT: WHY IS IT RELEVANT?**

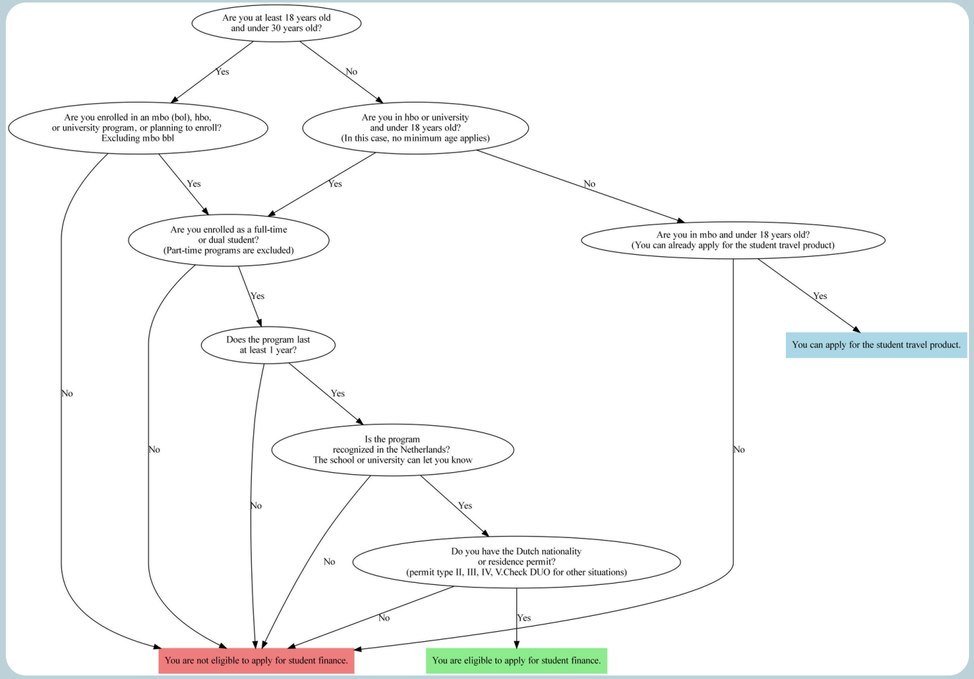
Approaches

- LLAMA 3.2: A COMPACT AND EFFICIENT TRANSFORMER MODEL OPTIMIZED FOR FAST INFERENCE. IT BALANCES PERFORMANCE AND SCALABILITY WITH LOW MEMORY USAGE, MAKING IT IDEAL FOR RESOURCE-CONSTRAINED ENVIRONMENTS.
- QWEN 2.5: A NEXT-GENERATION TRANSFORMER WITH STRONG MULTILINGUAL CAPABILITIES. IT FEATURES RETRIEVAL-AUGMENTED GENERATION (RAG) AND ENHANCED CONTEXTUAL AWARENESS FOR BETTER ACCURACY IN DIVERSE TASKS.
- DEEPSEEK-R1: A HIGH-CAPACITY MODEL WITH ADVANCED REASONING AND LONG-CONTEXT UNDERSTANDING. IT LEVERAGES DEEP ARCHITECTURAL OPTIMIZATIONS FOR IMPROVED EFFICIENCY IN COMPLEX PROBLEM-SOLVING.



Task-specific decision tree

AFTER A LEGAL ANALYSIS OF DUO STUDENT FINANCE REGULATIONS, VALIDATED BY A DOMAIN EXPERT, WE TRANSLATED THE PROBLEM INTO A DECISION TREE. ENCODED IN JSON FOR SEAMLESS LLM INTEGRATION, IT ENABLES SYSTEMATIC REASONING. ALSO, ANALYZING THE MODEL'S TREE TRAVERSAL PROVIDES DEEPER INSIGHTS INTO ITS DECISION-MAKING, ENHANCING INTERPRETABILITY AND TRANSPARENCY.

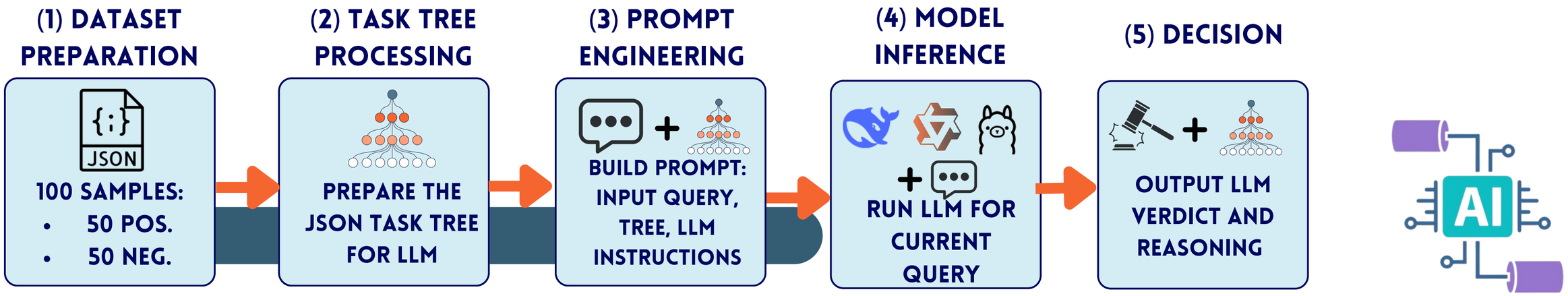


WE CONSTRUCTED A DATASET OF 20 REAL "DUO STUDENT FINANCE" CASES WITH GROUND TRUTH LABELS, EXPANDING IT TO 100 SAMPLES USING GENERATIVE AI. THE DATASET INCLUDES 50 POSITIVE AND 50 NEGATIVE CASES, STORED IN JSON, WITH DESCRIPTIONS WRITTEN IN NATURAL LANGUAGE TO CHALLENGE MODEL REASONING.


```
{
    "id":           92,
    "description": "Thomas is 30 years old and pursuing a
                    recognized bachelor 2019 degree in marketing
                    at an HBO institution in the Netherlands. The
                    program lasts three years, and he holds a
                    type. V residence permit.",
    "decision":     "NotEligible"
}
```

Pipeline architecture

THE PIPELINE PROCESSES A USER QUERY USING AN LLM (E.G., LLAMA 3.2), INCORPORATING A TASK-SPECIFIC JSON DECISION TREE TO ENHANCE MODEL REASONING, TO FINALLY PRODUCE THE VERDICT WITH DECISION PATH. THE EXPERIMENTS ARE COMPUTED USING OLLAMA.



TO REFINE PROMPT ENGINEERING, WE APPLY SEVERAL TECHNIQUES:

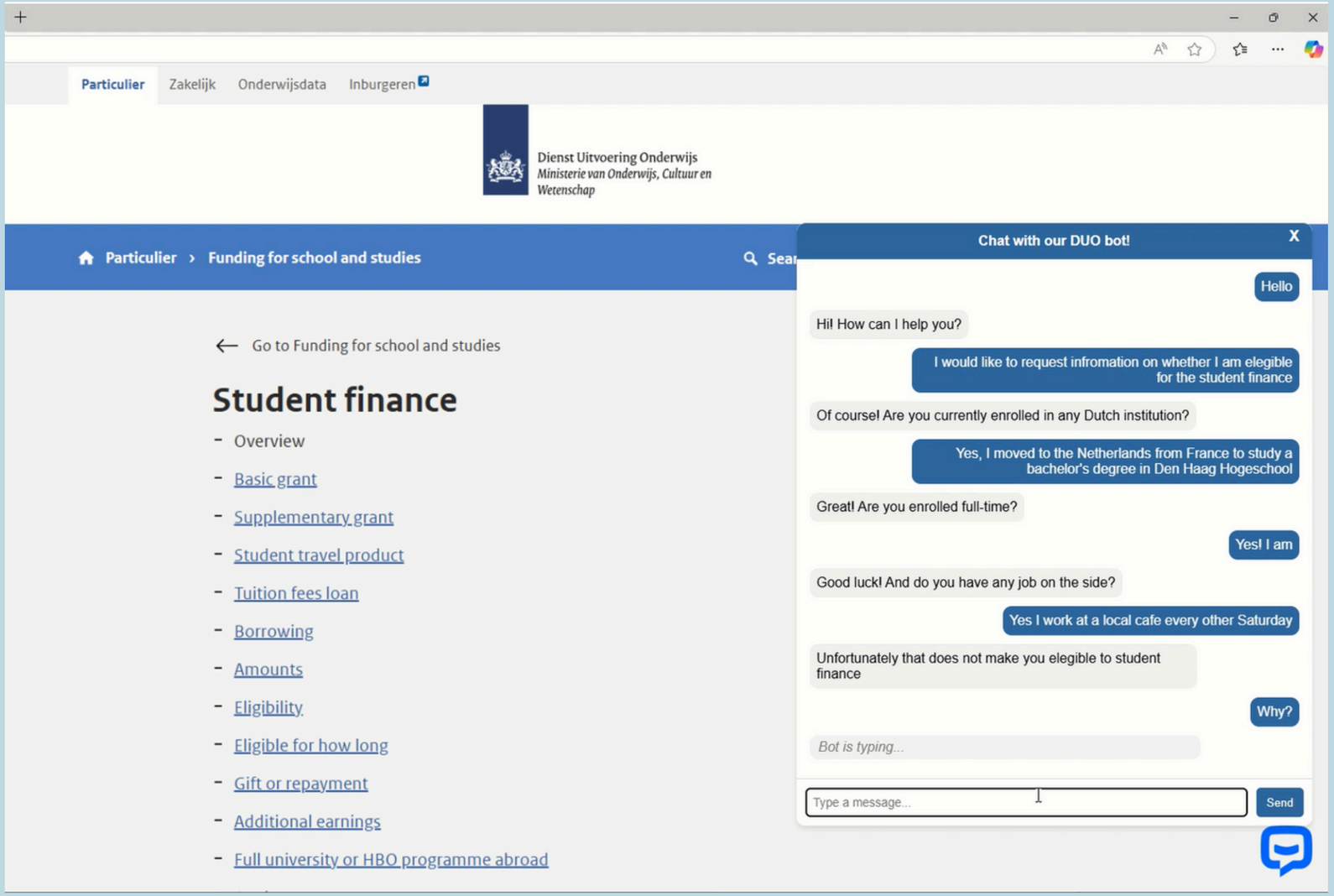
- **STRUCTURED QUERY FORMULATION – COMBINING USER QUERY, DECISION TREE, AND LLM INSTRUCTIONS FOR LOGICAL REASONING.**
- **ENFORCED JSON OUTPUT – STANDARDIZING PREDICTION, IMPACT NODE, AND REASONING TO MAINTAIN STRUCTURED RESULTS.**

Results

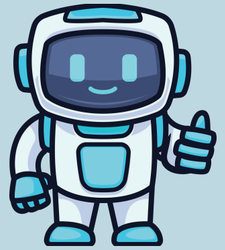
BASED ON THE EXPERIMENTAL RESULTS, WE DETERMINE THAT THE DEEPSEEK-R1 MODEL INTEGRATED WITH THE TASK DECISION TREE IS THE OPTIMAL CHOICE FOR THE FINAL VERSION OF THE JUS-TREE-AI PIPELINE.

Model	Parameters (B)	Temperature (0-1)	Accuracy (model-only)	Accuracy (model+tree)
Llama 3.2 [1]	3.2	0.8	0.53	0.58
Qwen 2.5 [2]	1.5	0.8	0.57	0.51
DeepSeek-R1 [3]	8.0	0.8	0.68	**0.86**

User interface



- INTEGRATED INTERFACE WITH THE DUO WEBSITE
- ALLOWS FOR A LIVE CHAT CONVERSATION WITH OUR DEVELOPED MODEL
- NOT ONLY DOES IT PROVIDE INFORMATION ABOUT SPECIFIC CASES, IT IS ALSO ABLE TO REASON AND GIVE THE APPROPRIATE RATIONALE BEHIND A DECISION



Future research

- Scalability to complex domains
- Dependence on ground truth traversal
- Model dependency and deneralizability
- Interpretability beyond traversals
- Potential for hallucination

- Automated graph construction
- Evaluating larger and more diverese datasets
- Real-world deployment and user-feedback
- Integrating feedback mechanisms

UNIVERSITEIT VAN AMSTERDAM

Thank you!

Any Questions?

Group D1: Filip Cierkosz, Gabriela Torres, Jobeal Solomon and Eduard Cruset

31-01-2025