Ironhack RoboReviews

LLM R&D Project Results

Katharina Krux

After thorough R&D, we propose a new AI-driven product for reputation management

Management Summary

- 5 days of **thorough R&D** by our **AI expert team**
- Proposed product consists of **three modules**:
  - **Sentiment analysis** of customer reviews
  - **Grouping** of top-rated **products**
  - **Automated blog post** writing
- Many **use-cases**:
  - Designed to **automate reputation management**
  - Also useful for **competitor analysis**
  - **Modules** can be used **stand-alone** or **in combination** for broader target audience
- **Good results** already achieved, but **further optimisation** necessary

➡ The research team is <u>seeking approval</u> to develop the product further
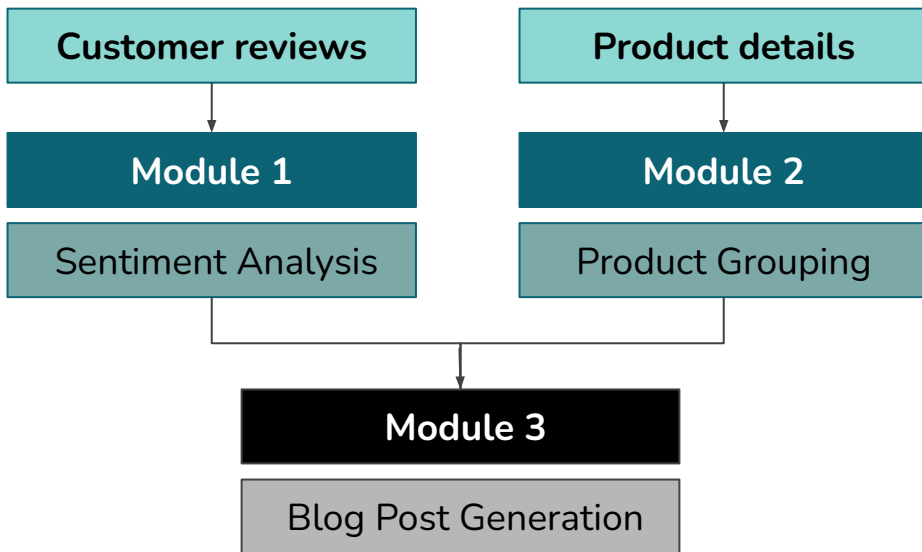
Overview and Business Case

Three modules: sentiment analysis, product categorisation and automated blog writing

Overview of the modules

Customer reviews	**Product details**
Module 1	**Module 2**
Sentiment Analysis	Product Grouping

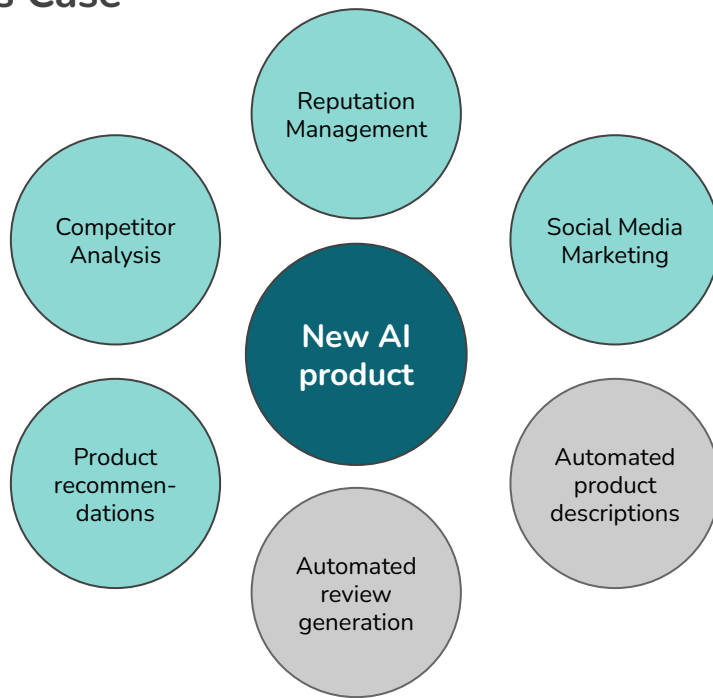**Module 3**

Blog Post Generation

Explanation:

- **Module 1** uses **customer reviews** and analyses **sentiment**

- **Module 2** uses **product data** to **group products** into **pre-defined** categories

- **Module 3** uses **output** of **both** Module 1 and 2 to generate **blog posts** of **top products**

- Can be **sold** as **complete package** or **separately**

Online reputation management growing in demand, AI can automate this process

Business Case

Reputation Management

Competitor Analysis

Social Media Marketing

New AI product

Product recommen-dations

Automated product descriptions

Automated review generation

Legend:

Already working

Potential new variations

Explanation:

- **Number** of **online businesses** and **online presence** of **businesses** is **growing rapidly**

- Services like **SEO**, **online marketing**, **reputation management** and **social media marketing** are **in demand**

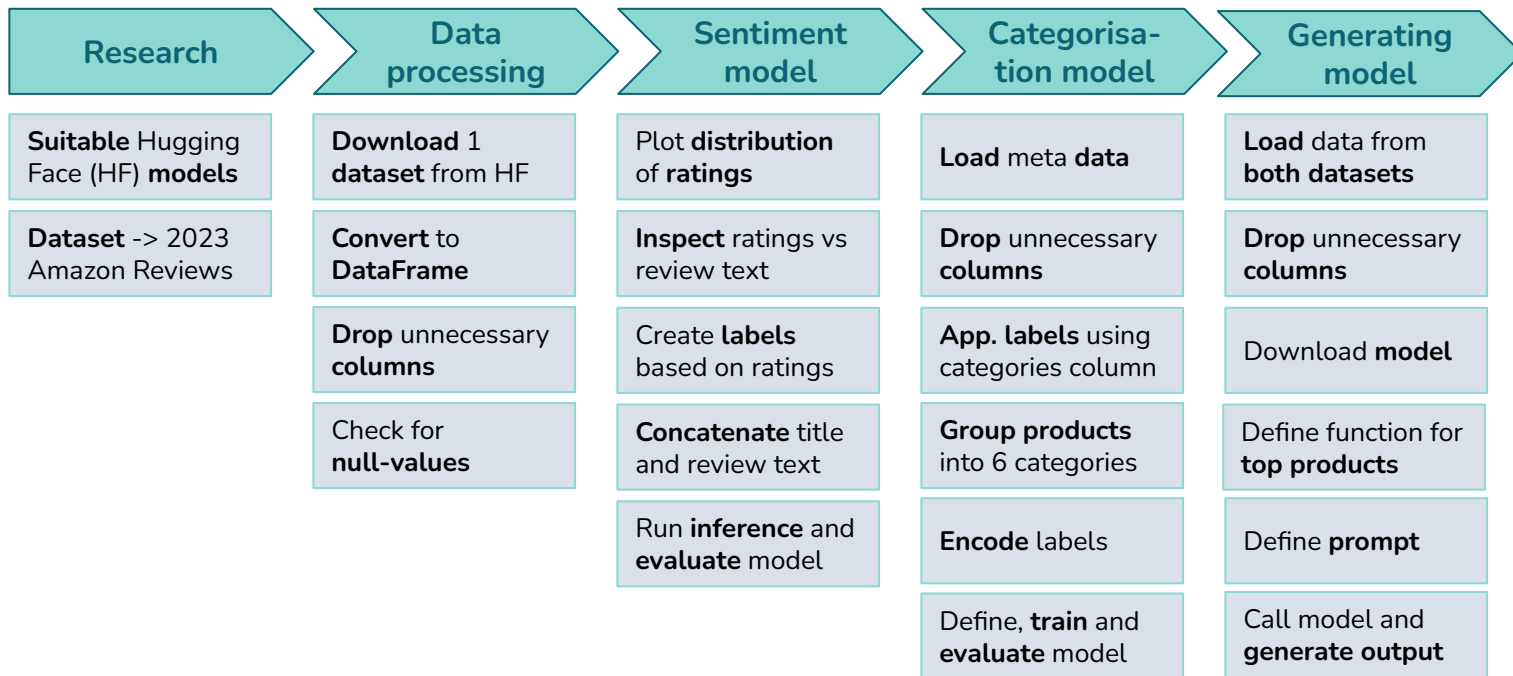- We aim to **create a product** that can **cater** to **all these needs**, by **leveraging AI/ LLMs**

Technical Aspects and Performance

Different parts of data used in each model, output of one model used as input for next

Methods and Workflow

Research	Data processing	Sentiment model	Categorisation model	Generating model
Suitable Hugging Face (HF) **models**	**Download** 1 **dataset** from HF	Plot **distribution** of **ratings**	**Load** meta **data**	**Load** data from **both datasets**
Dataset -> 2023 Amazon Reviews	**Convert** to **DataFrame**	**Inspect** ratings vs review text	**Drop** unnecessary **columns**	**Drop** unnecessary **columns**
	Drop unnecessary **columns**	Create **labels** based on ratings	**App. labels** using categories column	Download **model**
	Check for **null-values**	**Concatenate** title and review text	**Group products** into 6 categories	Define function for **top products**
		Run **inference** and **evaluate** model	**Encode** labels	Define **prompt**
			Define, **train** and **evaluate** model	Call model and **generate output**

Reputation management with Module 1 through sentiment analysis

Overview - Module 1

- Module 1 uses a **multilingual LLM** to **classify customer reviews** into one of **three categories**:
  - Positive
  - Neutral
  - Negative
- This can **help companies** by **screening reviews** to...
  - Identify **unhappy customers**
  - **Avoid/ remove bad reviews** off their page shortly after being posted
  - Use customer feedback to **improve products and services**
  - Identify **close competitors**
  - Identify **trending products** based on their reviews

Model: multilingual DistilBert Sentiment Student

Model already performing well, further optimisation possible for better accuracy

Evaluation - Model 1

- **Long run-time** of the model impacted opportunity to fine-tune
- Model **cannot be re-trained easily**, evaluation was **inference-only**
- **Accuracy** score based on **pre-defined labels** based on ratings
- **80.26% accuracy** is good, but **room for improvement**
- Confusion matrix shows **incorrect guesses** are **spread** across all categories
- **Visual inspection** of **incorrect guesses** not helpful

Inference results:
- Total test size: ~470.000 reviews
- Incorrect guesses: 92.388 items
- Accuracy: 80%

Confusion Matrix:
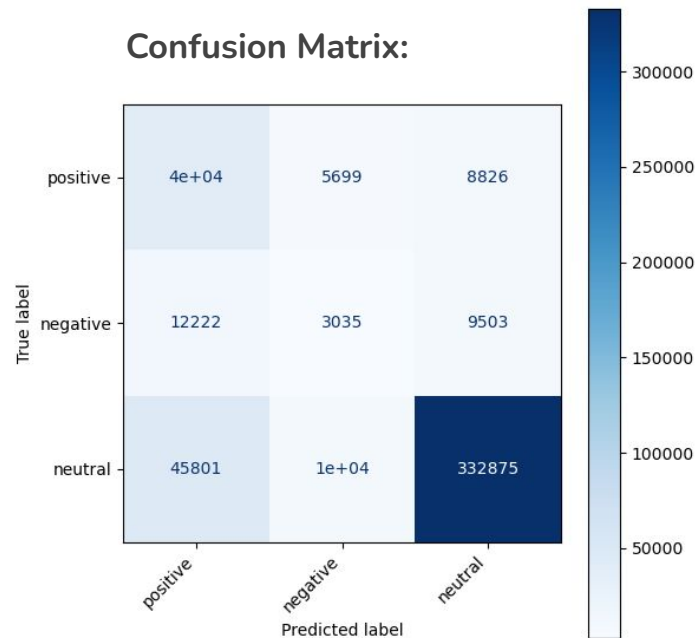
Categorisation of products into pre-defined groups with Module 2

Overview - Module 2

- Module 2 uses a **domain classifier LLM** to **categorise products** into one of **six categories**:
    - Dishwashers, washers and dryers
    - Microwaves, ovens and grills
    - Parts and accessories
    - Refrigerators, freezers and ice makers
    - Small appliances
    - Mixed
- This can **help companies** to…
    - Group top-rated products into the corresponding category
    - In combination with module 1:
        - Identify **trending products** from other websites based on the defined categories

Model: Domain Classifier by Nvidia

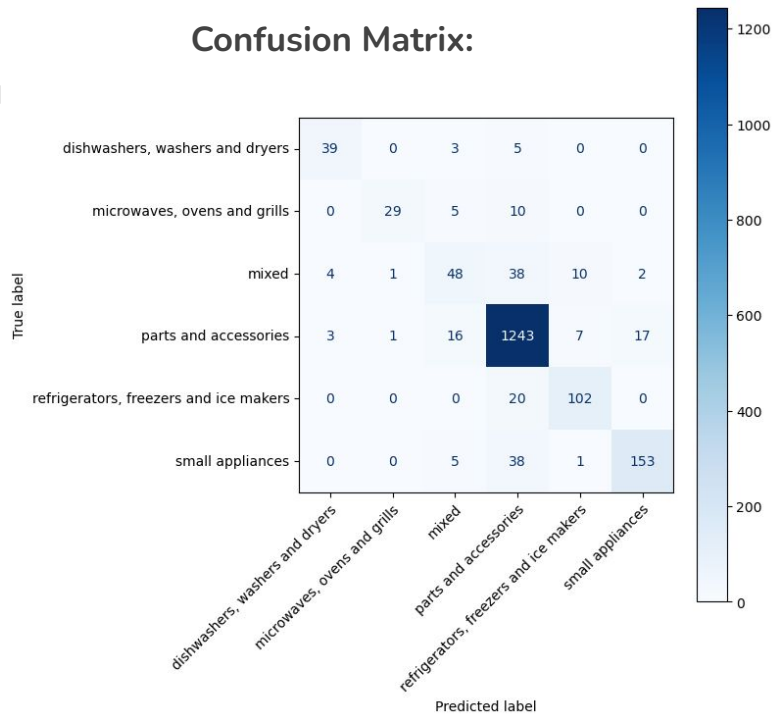# Due to time constraints, limited training took place (only 10% of data and 3 epochs)

Evaluation - Model 2

- **Long run-time** of the model impacted opportunity train and fine-tune
- Model had to be **re-trained** from **26 to** just **6** categories
- **Accuracy** score based on **pre-defined labels**, but **categories** as well as **product titles** in dataset were **not always clear**
- **Room for improvement** with **better pre-processing** and **longer training**
- Confusion matrix shows few **incorrect guesses**, specifically between **parts & accessories**, **small appliances** and **mixed**

Training details:
- Total training size: ~5.000 products (10%)
- Total testing size: ~ 1.000 products (10%)
- Training runtime: 3 epochs

Confusion Matrix:

| True label \ Predicted label	dishwashers, washers and dryers	microwaves, ovens and grills	mixed	parts and accessories	refrigerators, freezers and ice makers	small appliances
dishwashers, washers and dryers	39	0	3	5	0	0
microwaves, ovens and grills	0	29	5	10	0	0
mixed	4	1	48	38	10	2
parts and accessories	3	1	16	1243	7	17
refrigerators, freezers and ice makers	0	0	0	20	102	0
small appliances	0	0	5	38	1	153

Accuracy of 89.66% achieved after three epochs - more fine-tuning necessary

Evaluation - Model 2 (continued)
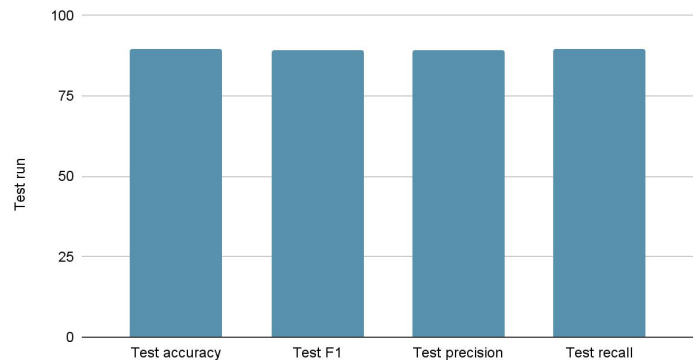
Training Results:

Training - Evaluation metrics

Epoch 1 ■ Epoch 2 ■ Epoch 3

Eval. accuracy | Eval. F1 | Eval. precision | Eval. recall
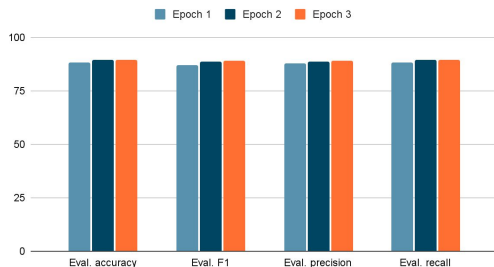
Training - Evaluation loss

Epoch 1 | Epoch 2 | Epoch 3
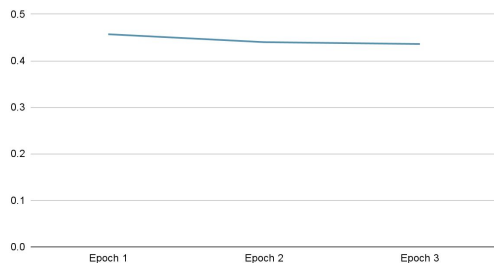
Testing Results:

Test run - evaluation metrics

Test accuracy | Test F1 | Test precision | Test recall

Metric	Accuracy	F1	Precision	Recall
Test Score (%)	89.66	89.22	89.20	89.66

AI automated Blog Post Generation for social media marketing with Module 3

Overview - Module 3

- Module 3 uses a **generative LLM** to create a **blog post** describing **top products**:
    - Summary of **product features**, based on **product information**
    - Summary of **pros**, based on **positive reviews**
    - Summary of **cons**, based on **negative reviews**
- This can **help companies** or **social media influencers** to:
    - **Increase** their **social media presence**
    - Make blog post writing **more efficient**
    - Conduct **targeted product research** by **filtering** and **summarising** information about the **top products** for a **certain category**

Model: Amber (based on LLama)

Model 3 output still suboptimal, bigger model should be considered

Evaluation - Model 3

- Model 3 with the **most difficulties**
- **Different types** of **prompts** were tested:

Tested prompts	Outcomes
Simple prompt with l**ittle instructions**	• The model output **information based** on its **own knowledge** • Sometimes **unrelated** to **products** (eg. project brief from random GitHub)
Large prompt with **example** and bigger context length	• The model output **information** directly **from the example** • Or it **gave information** for **similar products** (washing machines)
Broken down prompt, that **divides** the **task** into **three sections** and **combines outputs** in a final article	• Output **relates** to the **products given** • However, results are in a **"copy-paste" format** instead of summarisation

- Using a **larger model** should help make outputs **more realistic**
- Further **prompt engineering** is also recommended

The text generation still needs a lot of work, output is copy-paste of product information

Product: Fluidmaster 12IM72 Braided Stainless Steel Ice Maker Connector Water Line with Dual 1/4-In. x 1/4-In. Female Compression Threads, 6 Ft. (72-In.) Length
Features Summary:
Specifications : ['INCLUDES: One 1/4 in. x (72 inches) stainless steel ice maker connector', "One Ice Maker Connections Assembly", 'Four Nickel-plated Brass Hex Nuts, Washers and Splicing Tape']
Warranty : ['Limited Lifetime']]
Product dimensions: , 72.0 x .154" (3/8")
Positive Reviews Summary:
See all 3,106 customer reviews on Amazon USA
Negative Reviews Summary:
Reviews: Doesn't last and will ruin a floor - Review from Home Depot Started leaking two years after installing .Starts Leaked at area just above nut part ,made in China ... One Star didnt replace my plastic icemaker tubing which crack... maybe its a peculiarity of the fridge, don't know.
Reviews: Fluidmaster 12IM74 Stainless Steel Ice Making Connector Kit with Dual Compression Fittings and Adapter - White (3-Pack) Review Started leaks after one year Installing in a refrigerator made by Kenmore, probably from poor design of the plastic faucet.
Reviews: Fluidmaster 12IM74 Stainless Steel Ice Making Connector Kit with Dual Compression Fittings and Adapter - White (3-Pack) Review Started leaks after one year Installing in a refrigerator made by Kenmore, probably from poor design of the plastic faucet.
Reviews: › Fluidmaster IM72 Stainless Steel Ice Making Connector Kit with Dual Compression Fittings and Adapter Review - Started leaks after one year Installing in a refrigerator made by Kenmore, probably from poor design of the plastic faucet.
Reviews: Fl
……
Product: Refrigerator Lock,Fridge Lock with Keys,Freezer Lock and Refrigerator Lock for Child Proof(Fridge Lock-Black 1Pack)
Features Summary:
4) [Please tell us] how would describe this product 'Refrigerator Lock(Fridge Lock-Black/Freezer Lock):' in detail? ['Describe your company:Our factory was established more than ten years, We mainly produce safety padlocks and locking products.' ,'We have many kinds of locks for choice:Door Security PadLock / Furniture Drawer Locker Safety Box DooPad Lock/Wardrobe Cabinet Safe Cupboard Slide File Cabinets GunSaferes 45L Fireproof Water-Resistant Carry Case Bag,etc.']
'Please tell us the size of this product: 'Refrigerator Lock(Fridge Lock):' Please provide images and video to illustrate your answer https://youtu[...]<jupyter_text>Step2 - Ask Users for their contact detailsThe following questions will be asked if a user wants to purchase the item. If they do not wish, you may choose whether or how many times this question should appear.<NAME>: 'What is your name?': ['Your Name', '', ''],<jupyter_code># Step 2 - Ask Users for their contact details
def askUserForContact(contact):   # A function that asks user to enter a value of type string.
        if (contact == "phone" or str("")==str
Positive Reviews Summary:
Reviews: Easy Installation This product is great! I'm so happy that i can protect my children from getting into all of our foods on our grocery list, and even items in the freezer like ice cream or other frozen snacks we have stored up there too…so far it's been easy to install as well. Easy Installation I liked how simple this was! The instructions were clear enough that I didn't need any help with installing them on my fridge and they worked very well for what i needed, which was keeping all of the kids from opening up our refrigerator everyday without permission or getting into anything we don' t want them to eat. They are quite easy as you can see in this video here: https :// v= 48QZyDj5BGg I would recommend these for anyone who has kids!
Reviews: A Must Have For Every Family These locks were exactly what we needed at home, and they work great. They are easy to install (you have to drill a small hole in the back of your fridge so that it can be screwed on) but once you get used t o putting them there is no problem with installing or taking off these locks again!
Reviews: A Must Have For Every Family These lock work great and are easy for me to install. I would recommend this product if anyone has a
Negative Reviews Summary:
Reviews: Easy. To install, I took apart an old keyless entry padlock that was very easy for me because I am handy with tools in general…and cut the wires coming out of this lock into two 3/8" holes drilled into the back wall near the top and bottom edges (you can use a hole saw, but drill carefully to avoid damage or burrs on your metal surface).
Reviews: It doesn't work. I have tried everything possible including cutting wires with scissors , it is not even sticky at all !!! Please fix this product . Doesn'T Work Sticking in the freezer lock didn't seem to help any, but i will keep trying
Reviews: It does nothing and doesn't stay on. I was excited that this worked with our new fridge because of reviews saying it did so we could put locks for small children without having a keyed door...but the one you sent didn'T work at all..we couldn;td even get to open it
Reviews: It is not sticky enough. I tried and they came off within minutes when my child pushed them with her hand! Please fix this product or send me another replacement as I really need a fridge lock that would last for years without coming apart so easily 😔I had to use scissors on the wires, but it doesn't

Conclusion

Models need to be fine-tuned further, GPU usage and bigger models recommended

Takeaways and next steps

- **Positives:**
  - **Good results** were achieved **overall**
  - **Broad spectrum** of **use cases**
  - Use of **LLMs** and other AI models **enable automatisation** of tasks and **handling of big data**
- **Negatives:**
  - **Time lost** due to **constant debugging**, as well as issues with **data and model loading**
  - **Training** and inference **very slow** on **local machine** -> could only train models with **small part** of the **data** available
  - **Text generation model** needed to be **quantized** due to memory constraints -> could be part of the reason for bad output
  - **Text** data **not very structured**/ consistent, which made tasks like categorisation and summarisation **difficult** for the models
- **Takeaways:**
  - Use of **GPU power** would have **increased training speed** -> had to move from Paperspace to local halfway through
  - Importance of **prompt engineering**
- **Next steps:**
  - More **data preprocessing**
  - Try to **fine-tune** the models or **try** with **bigger models** for better results
  - **Deployment**

GitHub Repo: https://github.com/Katharina-code/Iron_RoboReviews