Improving Patent Classification using AI-Generated Summaries

Naoya Yoshikawa, Ralf Krestel

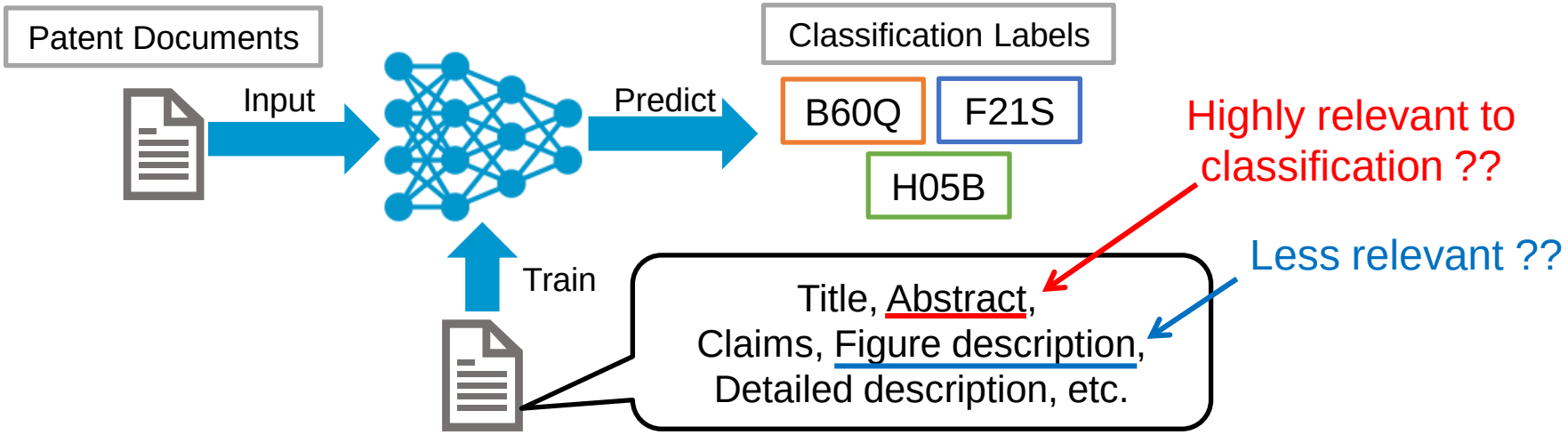
18.07.2024

Improving Patent Classification using AI-Generated Summaries

OR

Who Writes Better Patent Summaries: Humans or AI?

Patent Classification



Which Information to Use for Classification?

- Many previous studies used simple text extraction methods
  - e.g., abstract text, first 100 words of claims, etc.
- But is this good enough?
  - e.g. Patent for car exterior mirrors and approach lights

Class	Description of Class
B60Q	Arrangement of Signaling or Lighting Devices, …, for Vehicles in General
B60R	Vehicles, Vehicle Fittings, or Vehicle Parts, …

Abstract

An exterior mirror assembly including an attachment member for supporting an approach light. The attachment member interconnects a mirror housing to a vehicle and …

Claims

1. An exterior rear view mirror assembly comprising: a housing; a reflective mirror supported by the housing and arranged in a rearwardly facing direction; …

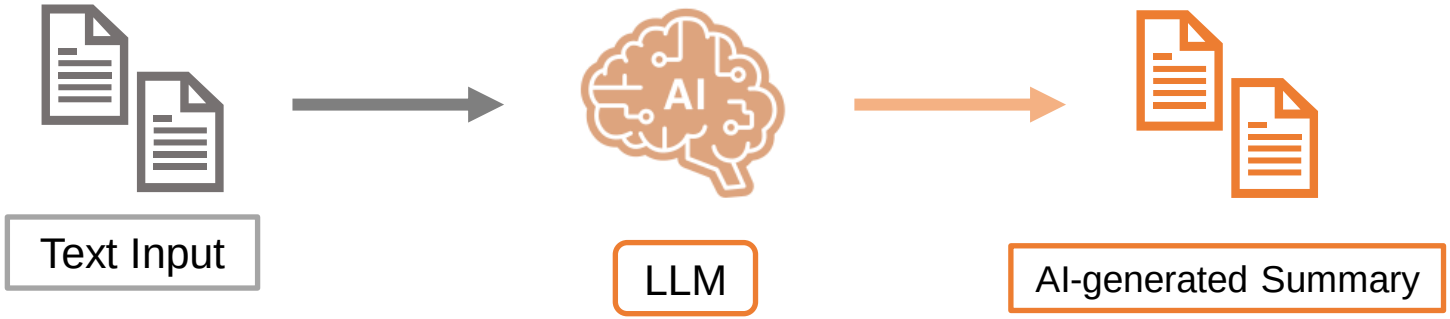
Detailed Description

DESCRIPTION OF THE PREFERRED EMBODIMENTS Referring now to the figures, and in particular FIGS. 1–3, a mirror assembly10 includes a housing12 which houses …

Cited: https://www.subarupartsdeal.com/parts/subaru-accessories_ext_auto_dim_mir_bsd_deal-j201ssg601.html

Large Language Models (LLMs) to the Rescue

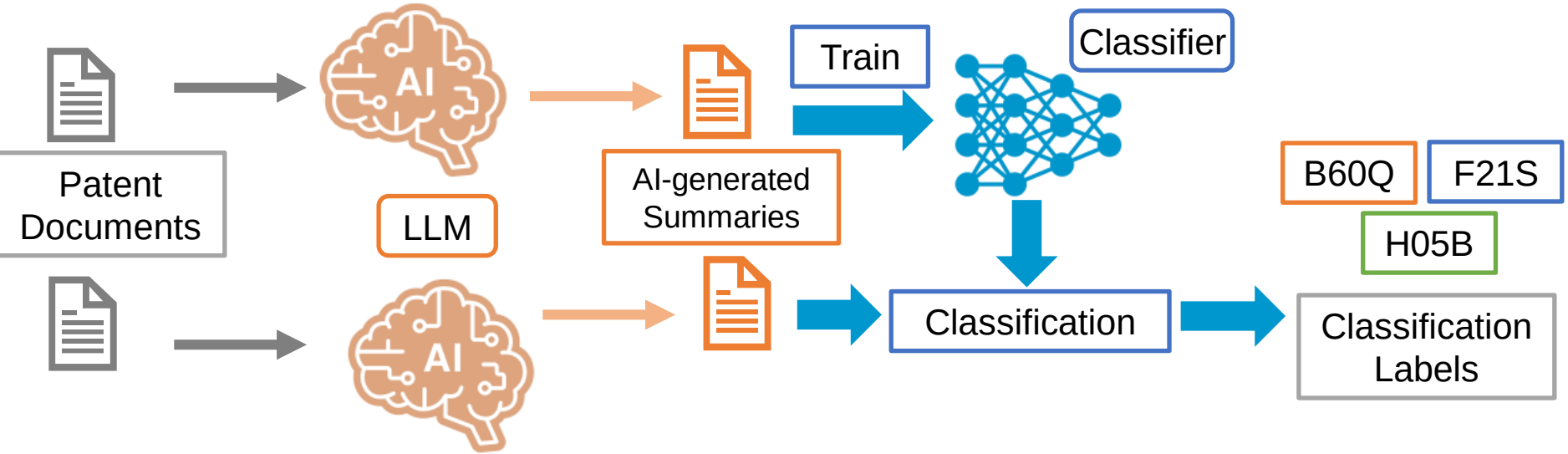
- LLMs have demonstrated high performance in various NLP tasks
- ChatGPT exhibits exceptional summarization capabilities (Yang et al. [1])
- LLMs have the potential to identify classification-effective information from patent documents



[1] X. Yang, Y. Li, X. Zhang, H. Chen, W. Cheng, Exploring the limits of chatgpt for query or aspect-based text summarization, 2023.

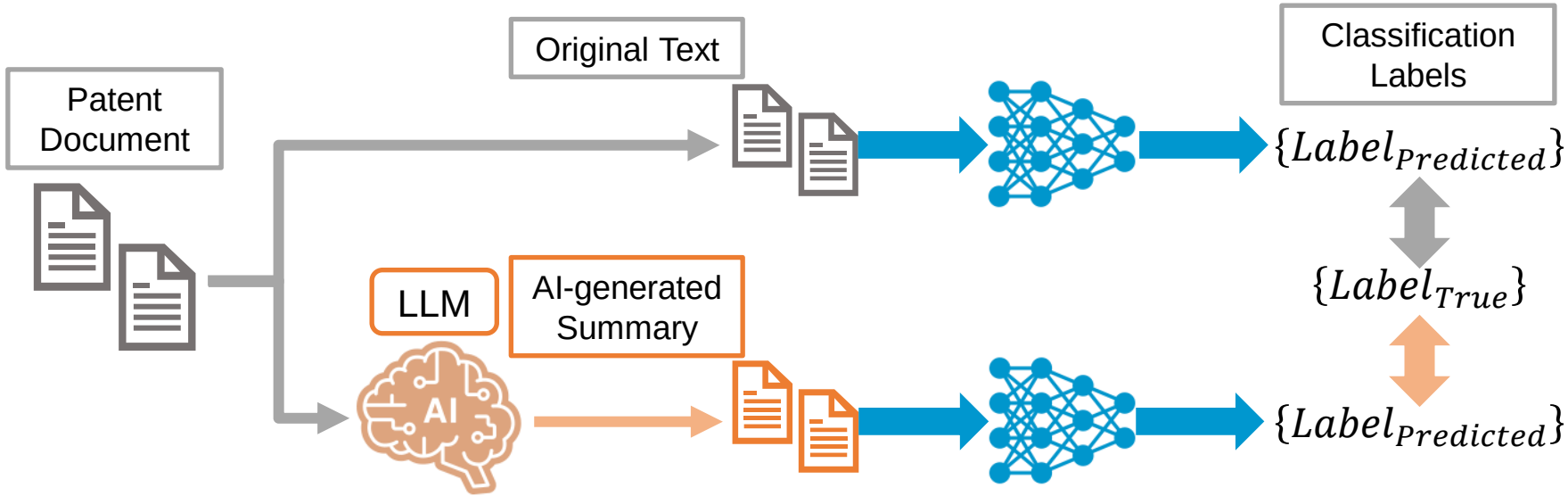
Objective

- Verify whether the performance of automated patent classification can be improved by using AI-generated summaries



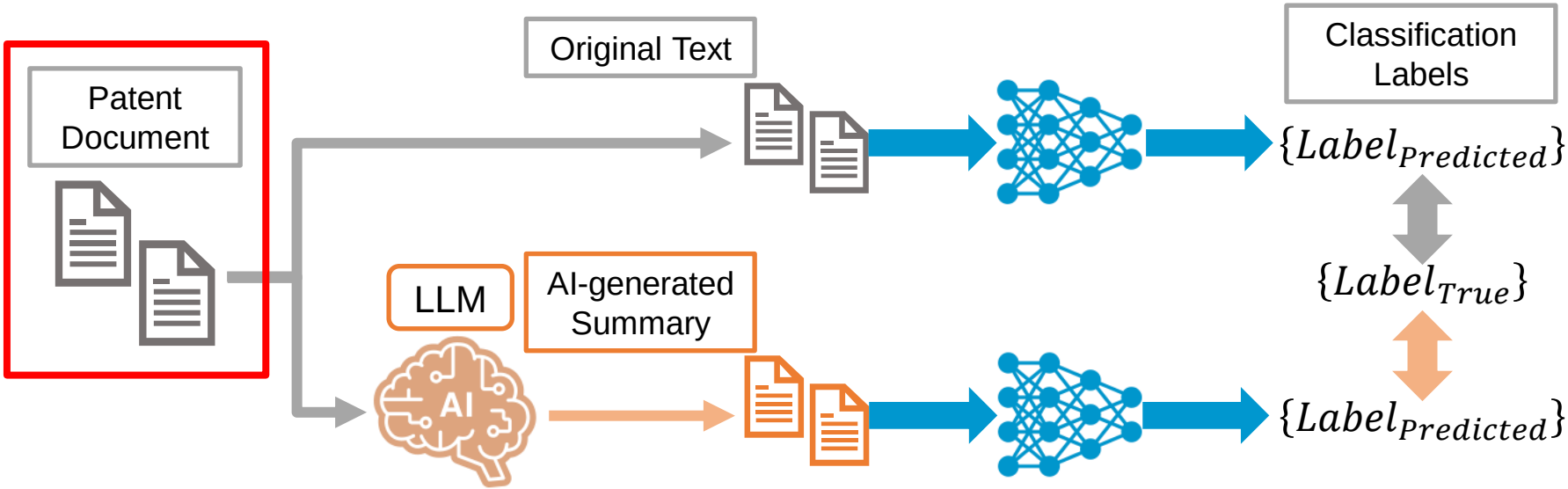
Experimental Setup

- Datasets
- Summary Generation
- Patent Classification



Experimental Setup

- Datasets
- Summary Generation
- Patent Classification



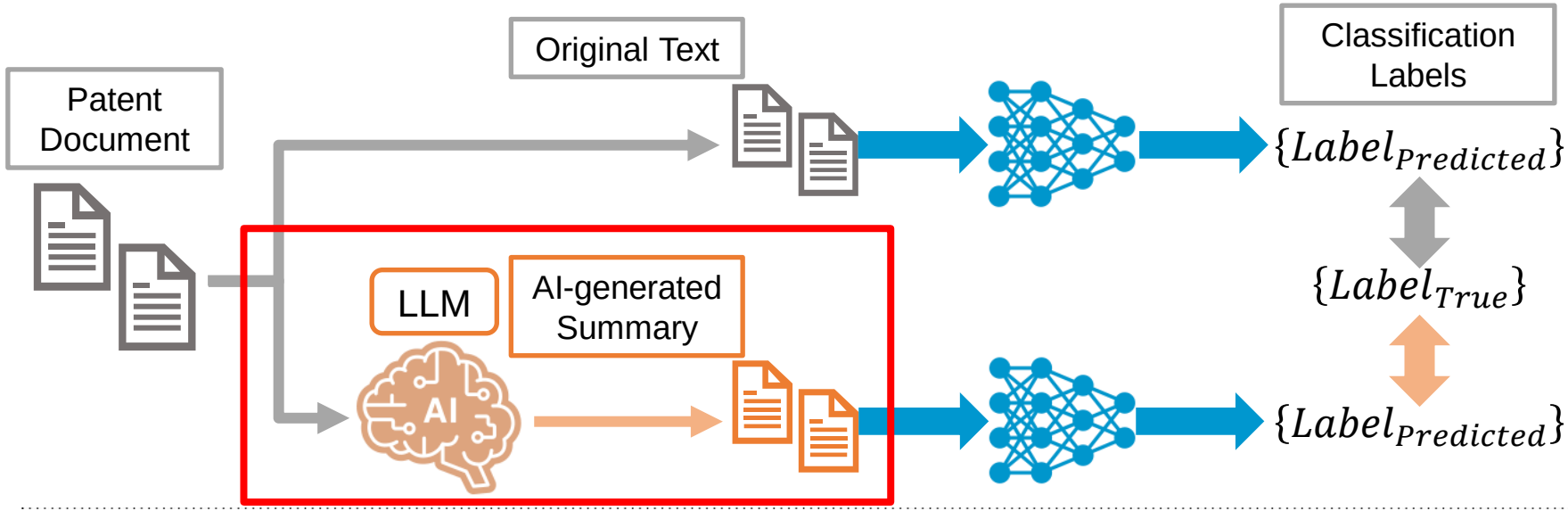
Datasets

- USPTO-70k dataset
  - USPTO patent documents from 2006 to 2019
- Subset of CLEF-IP 2011 dataset
  - EPO English patent documents from 2000 to 2009
  - Extracted from CLEF-IP 2011 dataset consisting of over 2.6 million patent documants

	Training	Validation	Test
USPTO-70k			
Subclass-level Multi-label			
# of documents	50,625	10,000	10,000
Avg. labels per patent	1.98	2.25	2.32
Total labels	630	573	585
Subgroup-level Multi-class			
# of documents	30,251	4,784	4,288
Total labels	1282	952	860
CLEF-IP 2011 subset			
Subclass-level Multi-label			
# of documents	50,000	10,000	1,000
Avg. labels per patent	2.00	2.01	1.87
Total labels	605	554	323

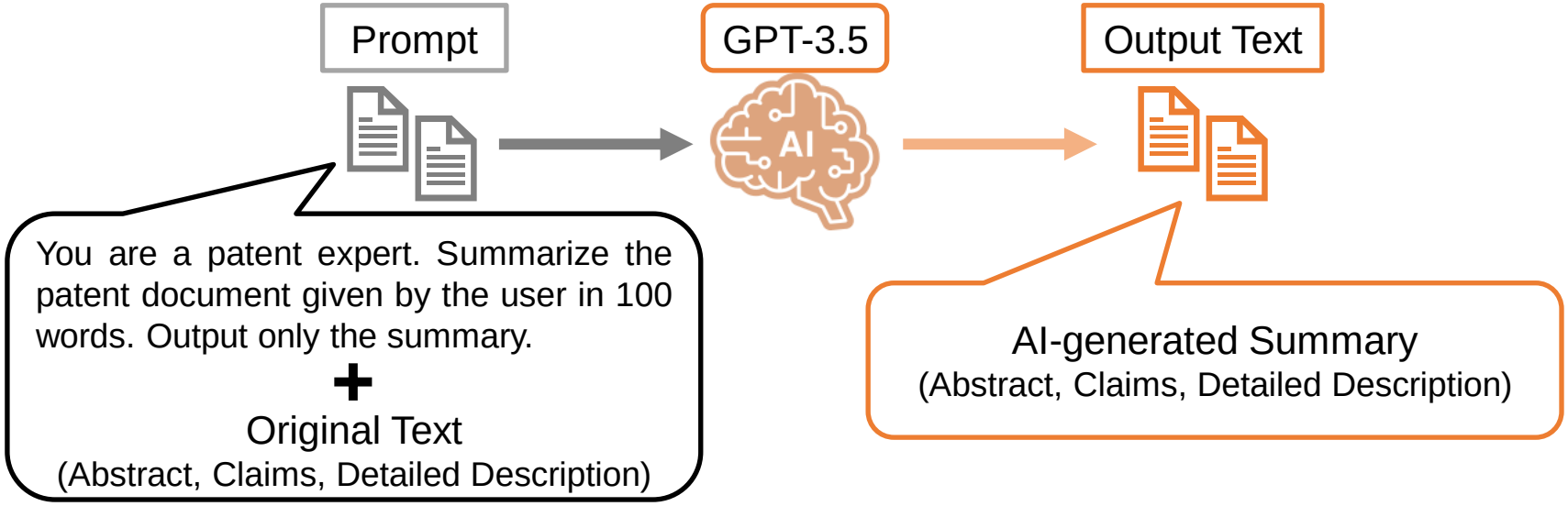
Experimental Setup

- Datasets
- Summary Generation
- Patent Classification



Summary Generation

- Model: OpenAI GPT-3.5 Turbo (gpt-3.5-turbo-0125)
  - Context Window: 16,385 tokens
  - Knowledge cutoff: September 2021



Summary Generation: Example 1

- Abstract (Original Text) covers main elements of invention
- Abstract (AI-generated Summary) retains key information from original text but expressed in plainer language

Abstract (Original Text)

An exterior mirror assembly including an attachment member for supporting an approach light. The attachment member interconnects a mirror housing to a vehicle and includes an opening for receiving a lens. Light projects through the lens from the attachment member in order to illuminate a predetermined area in proximity to the vehicle. A light source may be housed within the support member or, alternatively, a light source may be housed interior to the vehicle and a light path transport light from the light source to the lens for projection from the support member.

Abstract (AI-generated Summary)

The patent describes an exterior mirror assembly with an attachment member supporting an approach light. The attachment member connects the mirror housing to a vehicle and has an opening for a lens. Light passes through the lens to illuminate an area near the vehicle. The light source can be within the attachment member or inside the vehicle, with a light path transporting light from the source to the lens for projection.

Summary Generation: Example 2

- Claims (AI-generated Summary) specifically describe components of invention
- Detailed Description describes most comprehensive content of invention, such as its structure, function, and purpose

Claims (AI-generated Summary)

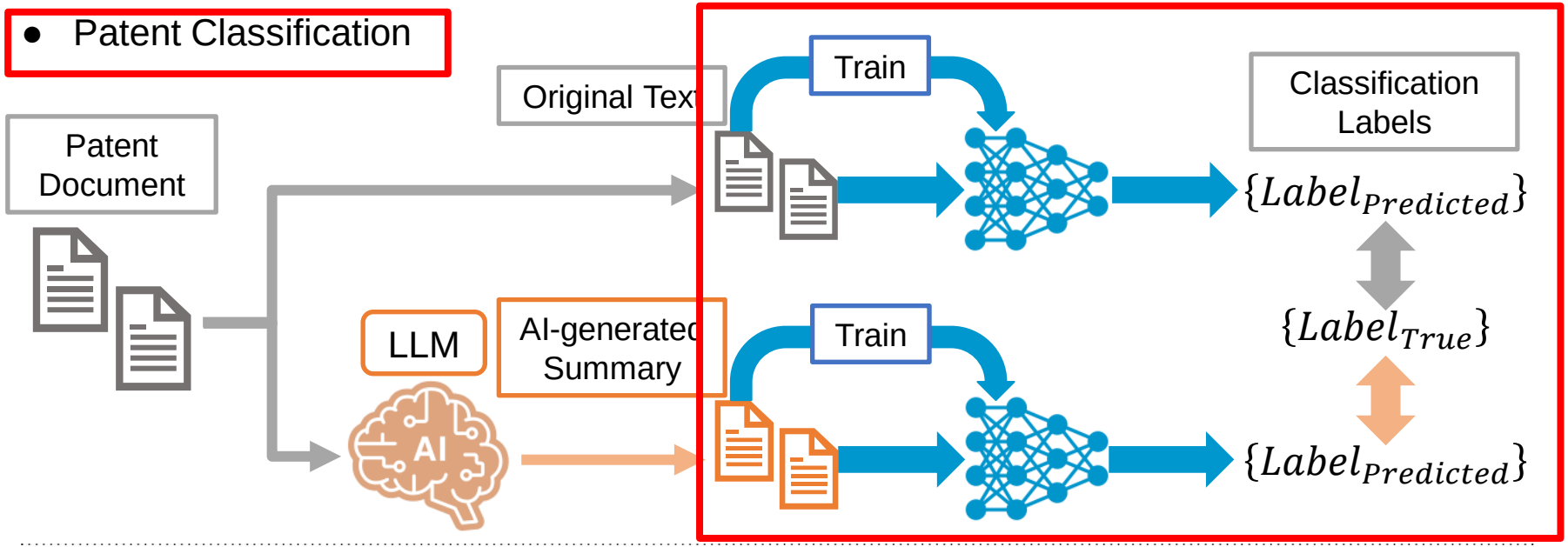
The patent document describes an exterior rear view mirror assembly for a motor vehicle. It includes a housing, a reflective mirror, a support member with a light transmitting lens for illuminating an area near the vehicle, and optional features like a light assembly with adjustable light source, reflective elements, environmental seal, sail for securing the support member, and a cover for the lens. The system can also include a timer, speed sensor, switch, and control circuit for activating the light source based on vehicle speed or various sensors.

Detailed Description (AI-generated Summary)

The patent document describes a mirror assembly for vehicles with an integrated approach light system. The mirror assembly includes a housing supporting a reflective element and an attachment member with a sail portion and a support arm portion. The approach light can be located in the support arm portion or sail portion, providing illumination for areas adjacent to the vehicle. Various configurations for the approach light placement, activation, and control are discussed, including adjustable reflectors, variable timeout functions, and integration with vehicle systems. The invention aims to improve visibility, safety, and serviceability of the approach light system in vehicle mirror assemblies.

Experimental Setup

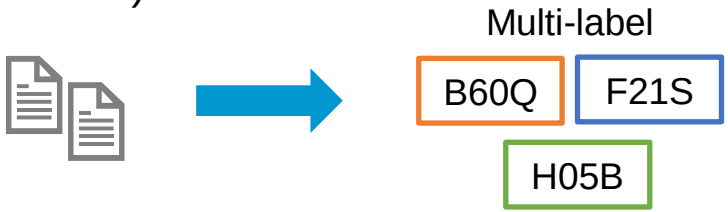
- Datasets
- Summary Generation
- Patent Classification



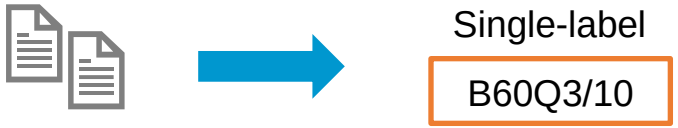
Patent Classification

- Fine-tuning RoBERTa model (roBERTa-base) to classification tasks

TASK 1: Subclass Classification
(all subclass labels)



TASK 2: Subgroup Classification
(only main subgroup label)



Hierarchical Structure of Patent Classification

Level	Label	Description
Section	B	Transporting
Class	B60	Vehicles in General
Subclass	B60Q	Arrangement of Signaling or Lighting Devices, …
Group	B60Q3	Arrangement of Lighting Devices for Vehicle Interiors
Subgroup	B60Q3/10	for Dashboards

Results: Subclass Classification

USPTO-70k

Text field	Precision	Recall	F1-score	Macro-F1
AB (OT)	53.6	52.8	53.2	21.4
CL (OT)	52.1	52.6	52.	21.3
DD (OT)	46.8	39.3	42.7	14.8
AB (AI)	53.9	52.8	53.4	21.4
CL (AI)	55.6	55.6	55.6	24.0
DD (AI)	**56.2**	**56.1**	**56.1**	**24.4**

CLEF-IP 2011

Text field	Precision	Recall	F1-score	Macro-F1
AB (OT)	59.5	57.6	58.5	18.7
CL (OT)	59.8	55.2	57.4	18.2
DD (OT)	60.3	60.3	60.3	20.9
AB (AI)	60.7	57.1	58.8	18.8
CL (AI)	62.1	59.6	60.8	20.2
DD (AI)	**63.0**	**63.3**	**63.1**	**22.8**

AB: Abstract, CL: Claims, DD: Detailed description, OT: Original text, AI: AI-generated summary

Baseline : Abstract (Original Text)

Results: Subgroup Classification

USPTO-70k

Text field	Accuracy	Macro-F1
AB (OT)	22.3	11.0
CL (OT)	21.9	11.1
DD (OT)	14.3	6.5
AB (AI)	22.0	11.3
CL (AI)	**23.9**	12.9
DD (AI)	23.5	**13.1**

AB: Abstract, CL: Claims, DD: Detailed description, OT: Original text, AI: AI-generated summary

Baseline : Abstract (Original Text)

Prompt Comparison

Prompt type	System role message
Patent	You are a patent expert. Summarize the patent document given by the user in 100 words. Output only the summary.
Simple	Summarize the document given by the user in 100 words. Output only the summary.
Elaborate	You are a patent expert. Summarize the patent document given by the user in 100 words focusing on an addition to the state of the art. "Addition to the state of the art" means the difference between the subject matter in a patent document and the collection of all technical subject matter that has already been placed within public knowledge. Output only the summary.

Results - Prompt Comparison

TASK 1: Multi-label Subclass-level Classification

Dataset	Text field	Prompt	Precision	Recall	F1-score	Macro-F1
USPTO-70k	DD (AI)	Patent	56.2	56.1	56.1	24.4
		Simple	56.0	56.1	56.0	24.8
		Elaborate	55.3	56.1	55.7*	24.7
CLEF-IP 2011	DD (AI)	Patent	63.0	63.3	63.1	22.8
		Simple	63.5	64.2*	63.9*	23.6*
		Elaborate	62.8	62.1	62.4	21.9

TASK 2: Multi-class Subgroup-level Classification

Dataset	Text field	Prompt	Accuracy	Macro-F1
USPTO-70k	DD (AI)	Patent	23.6	13.1
		Simple	23.5	12.9
		Elaborate	24.4*	13.4

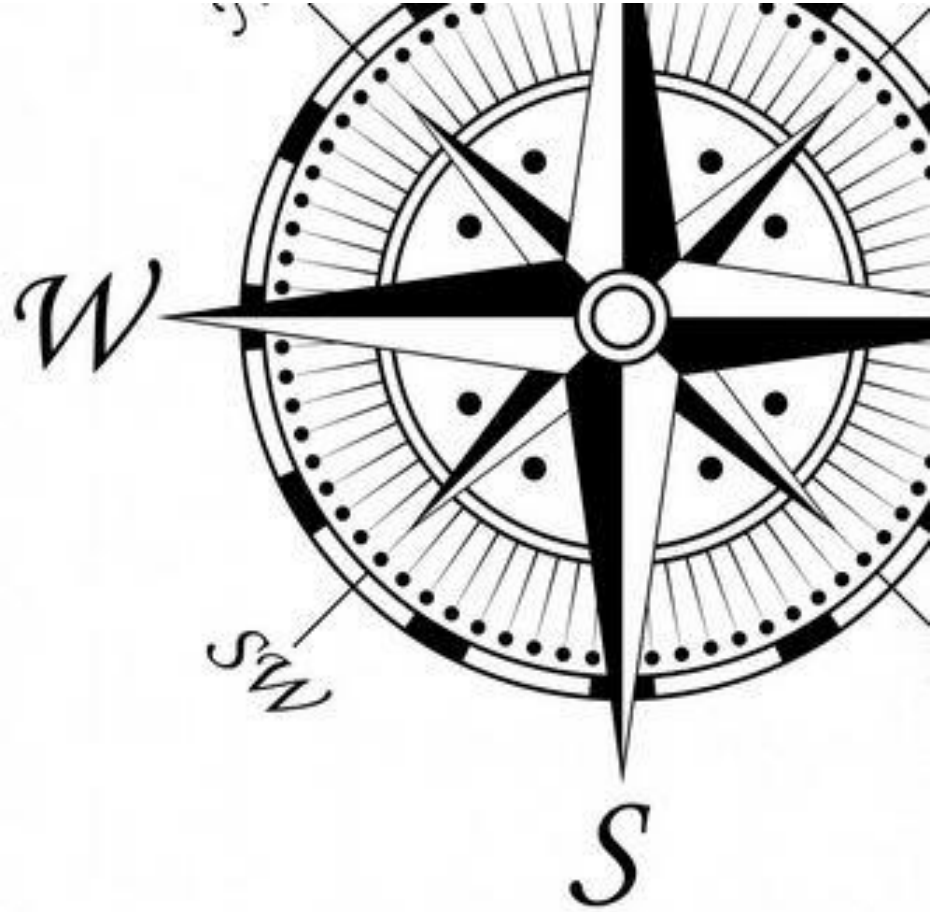
DD: Detailed description, AI: AI-generated summary, *p<0.05, **p<0.01, Baseline : “Patent” Prompt

Conclusion

- AI-generated summaries for patent classification outperform original abstract in both, subclass and subgroup classification
- LLMs seem to capture classification-relevant information better

- Who Writes Better Patent Summaries: Humans or AI?
  - If measured by suitability for classification, the answer is AI

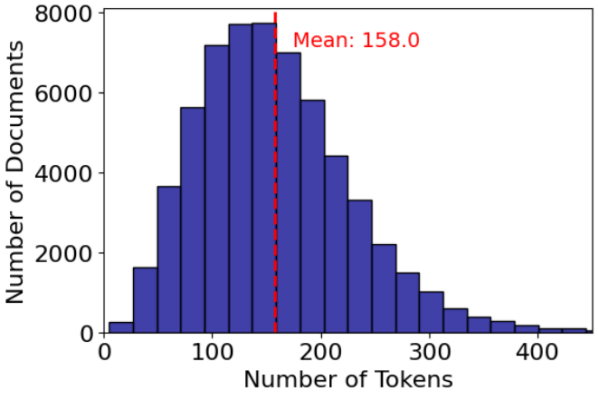
Supplementary materials



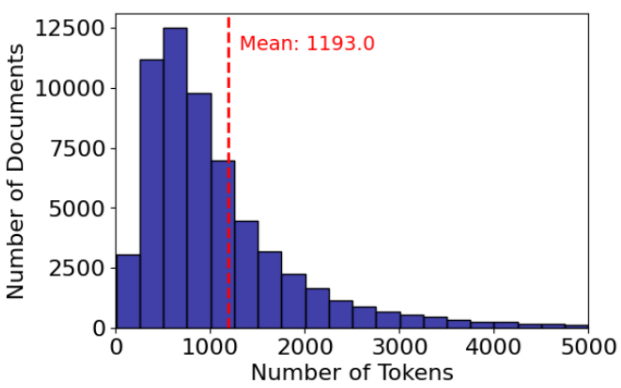
Datasets

- Token distributions for each different text field on CLEF-IP 2011 subset

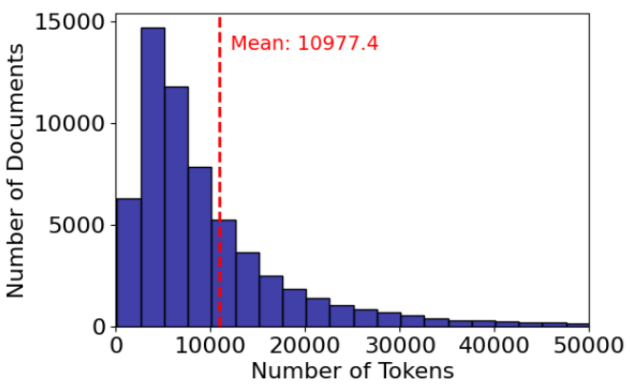
Abstract



Claims



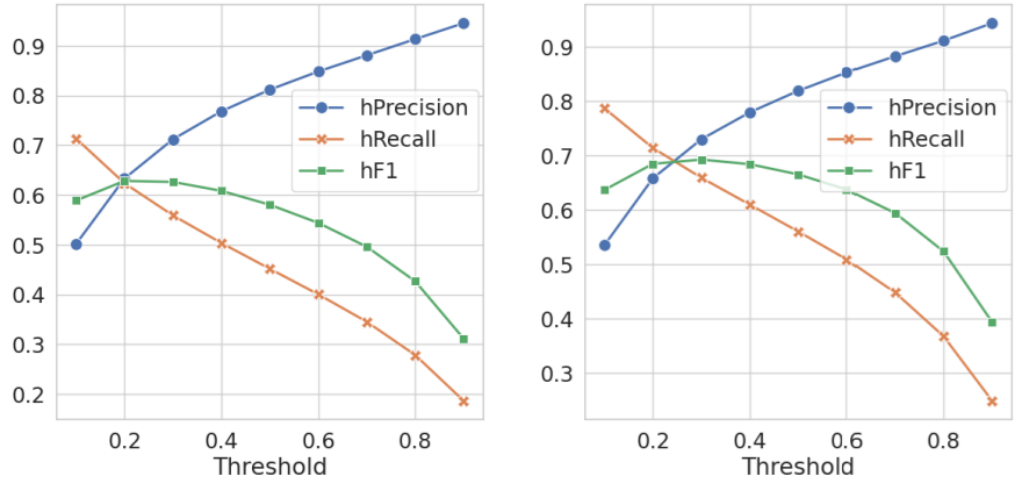
Detailed Description



RoBERTa Fine-tuning settings

Table 3: Hyperparameters for fine-tuning RoBERTa

Hyperparameter	Value
Batch size	64
Learning rate	5e-5
Number of epochs	15
Max sequence length	512
Dropout	0.2



(a) USPTO-70k dataset **(b) CLEF-IP 2011 subset**

Figure 2: Comparision of various threshold values for RoBERTa model (subclass-level classification, abstract original text)

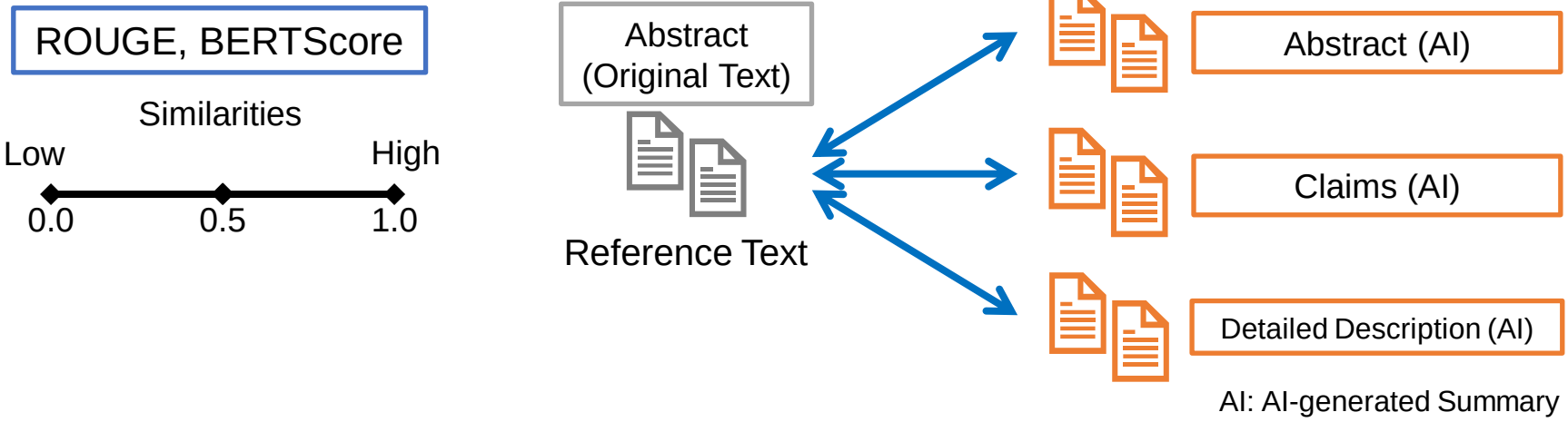
Text field	Text type	Text
Abstract	Original text	"An exterior mirror assembly including an attachment member for supporting an approach light. The attachment member interconnects a mirror housing to a vehicle and includes an opening for receiving a lens. Light projects through the lens from the attachment member in order to illuminate a predetermined area in proximity to the vehicle. A light source may be housed within the support member or, alternatively, a light source may be housed interior to the vehicle and a light path transport light from the light source to the lens for projection from the support member."
Abstract	AI-generated summary	"The patent describes an exterior mirror assembly with an attachment member supporting an approach light. The attachment member connects the mirror housing to a vehicle and has an opening for a lens. Light passes through the lens to illuminate an area near the vehicle. The light source can be within the attachment member or inside the vehicle, with a light path transporting light from the source to the lens for projection."

Text field	Text type	Text
Claims	Original text (first 100 words)	"1. An exterior rear view mirror assembly comprising: a housing; a reflective mirror supported by the housing and arranged in a rearwardly facing direction; a support member having an outboard end attached to the housing and an inboard end secured to an outer surface of a motor vehicle, the support member including an opening; and a light transmitting lens disposed in the opening of the support member, the lens enabling light to project from an interior of the support member to an exterior of the support member to illuminate a predetermined area in proximity to the motor vehicle. 2. The"
Claims	AI-generated summary	"The patent document describes an exterior rear view mirror assembly for a motor vehicle. It includes a housing, a reflective mirror, a support member with a light transmitting lens for illuminating an area near the vehicle, and optional features like a light assembly with adjustable light source, reflective elements, environmental seal, sail for securing the support member, and a cover for the lens. The system can also include a timer, speed sensor, switch, and control circuit for activating the light source based on vehicle speed or various sensors."

Text field	Text type	Text
Detailed Desc.	Original text (first 100 words)	"DESCRIPTION OF THE PREFERRED EMBODIMENTS Referring now to the figures, and in particularFIGS. 1–3, a mirror assembly10includes a housing12which houses and supports a reflective element14, such as mirror glass. Mirror assembly10of the present invention will be described with respect to an exterior driver side rear view mirror which enables the vehicle operator to view an area beside and rearward with respect to the vehicle. Housing12connects to an attachment member16including a sail portion18and an integral, laterally projecting support arm or bracket portion20. The interconnection between housing12and attachment member16enables housing12to pivot with respect to the support arm portion20. Particularly, housing12pivots rearward and"
Detailed Desc.	AI-generated summary	"The patent document describes a mirror assembly for vehicles with an integrated approach light system. The mirror assembly includes a housing supporting a reflective element and an attachment member with a sail portion and a support arm portion. The approach light can be located in the support arm portion or sail portion, providing illumination for areas adjacent to the vehicle. Various configurations for the approach light placement, activation, and control are discussed, including adjustable reflectors, variable timeout functions, and integration with vehicle systems. The invention aims to improve visibility, safety, and serviceability of the approach light system in vehicle mirror assemblies."

Summary Generation - Evaluation metrics

- ROUGE-1, ROUGE-L, BERTScore-F1
- Measures similarity between generated and reference text



Results - Summary Generation

- Both ROUGE and BERTscore decrease from abstract to claims to detailed description, with BERTscore showing less reduction than ROUGE
- GPT 3.5 Turbo model tends to generate contextually relevant summaries even for longer, more complex texts like claims and detailed descriptions

Dataset	Text field	Avg. # of words	ROUGE-1	ROUGE-L	BERTScore-F1
USPTO-70k	Abstract (AI)	59.4 ± 16.7	0.548	0.469	0.910
	Claims (AI)	86.5 ± 11.8	0.399	0.261	0.868
	Detailed Description (AI)	97.9 ± 12.8	0.372	0.236	0.862
CLEF-IP 2011	Abstract (AI)	63.4 ± 17.7	0.520	0.423	0.894
	Claims (AI)	85.3 ± 12.6	0.379	0.248	0.859
	Detailed Description (AI)	96.1 ± 7.4	0.357	0.225	0.855