



# PolicyPulse

**Dipika Kumar, Hamsini Sankaran,  
Kirti Prakash, Sreeram Ravinothala**



# MEET THE TEAM



**Hamsini  
Sankaran**



**Dipika  
Kumar**



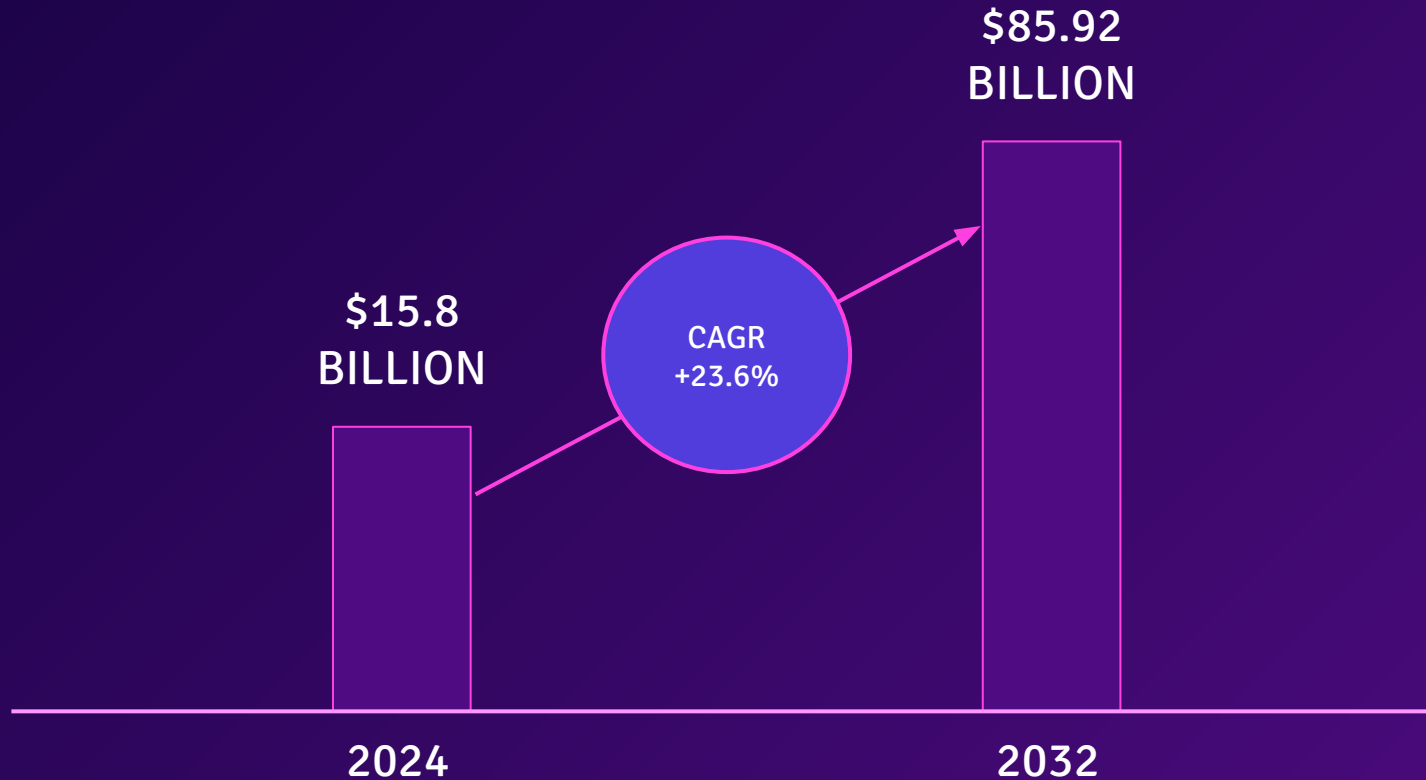
**Kirti  
Prakash**



**Sreeram  
Ravinoothala**

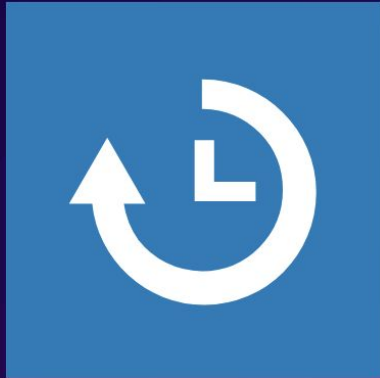
# Global RegTech Market

Driven by Increasing Regulatory Pressures

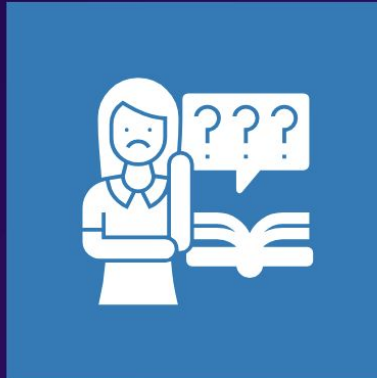


# Problem

*Fintechs & Law Firms invest significant resources and time towards privacy policy compliance with GDPR*



Manual Review



Legal Jargon



Hefty Fines

# Impact

## GDPR Non-Compliance Fines

**€20 million**

**OR**

**4 % of Global Annual  
Revenue**

**Uber**

**€290 Million in 2024**

**Meta**

**€1.2 Billion in 2023**

**amazon**

**€746 Million in 2021**

# Solution

**NLP-based application to streamline GDPR privacy policy compliance**

**Reduce Time**  
spent on policy review

**Avoid Risks**  
of fines & reputational damage

**Save Money**  
on billable attorney hours

**Early Detection**  
of GDPR violations

# Product

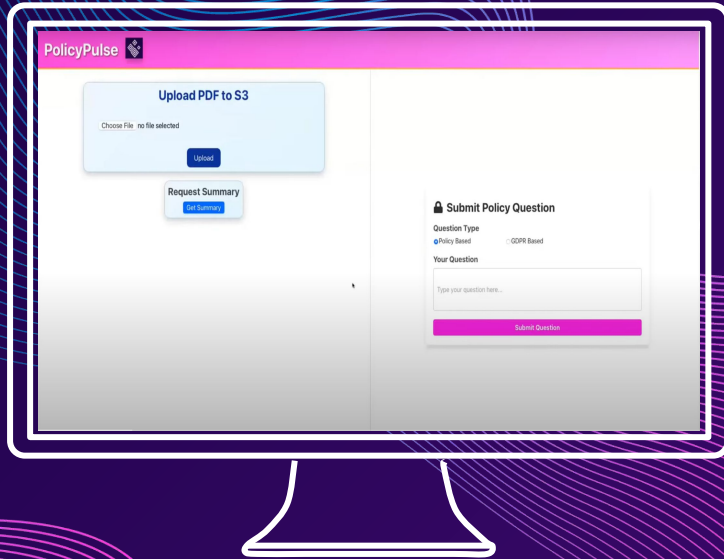


 **Question & Answering**

 **Summarization**

 **GDPR Checklist**

# DEMO



<https://youtu.be/meYclfv5bRw?feature=shared>





# Data and Model selection

## Key takeaways:

- Privacy Policy is complex
- Primary datasets low quality
- Ground Truth Generation needed, built with SMEs

## DATA

No target labels

OPP-115 Corpus

GDPR

CCPA

Privacy QA Dataset

LLM Generated Ground Truth ( SMEs )

## MODELS

Llama 2

Claude

Bart

Mistral 7b

Command R+

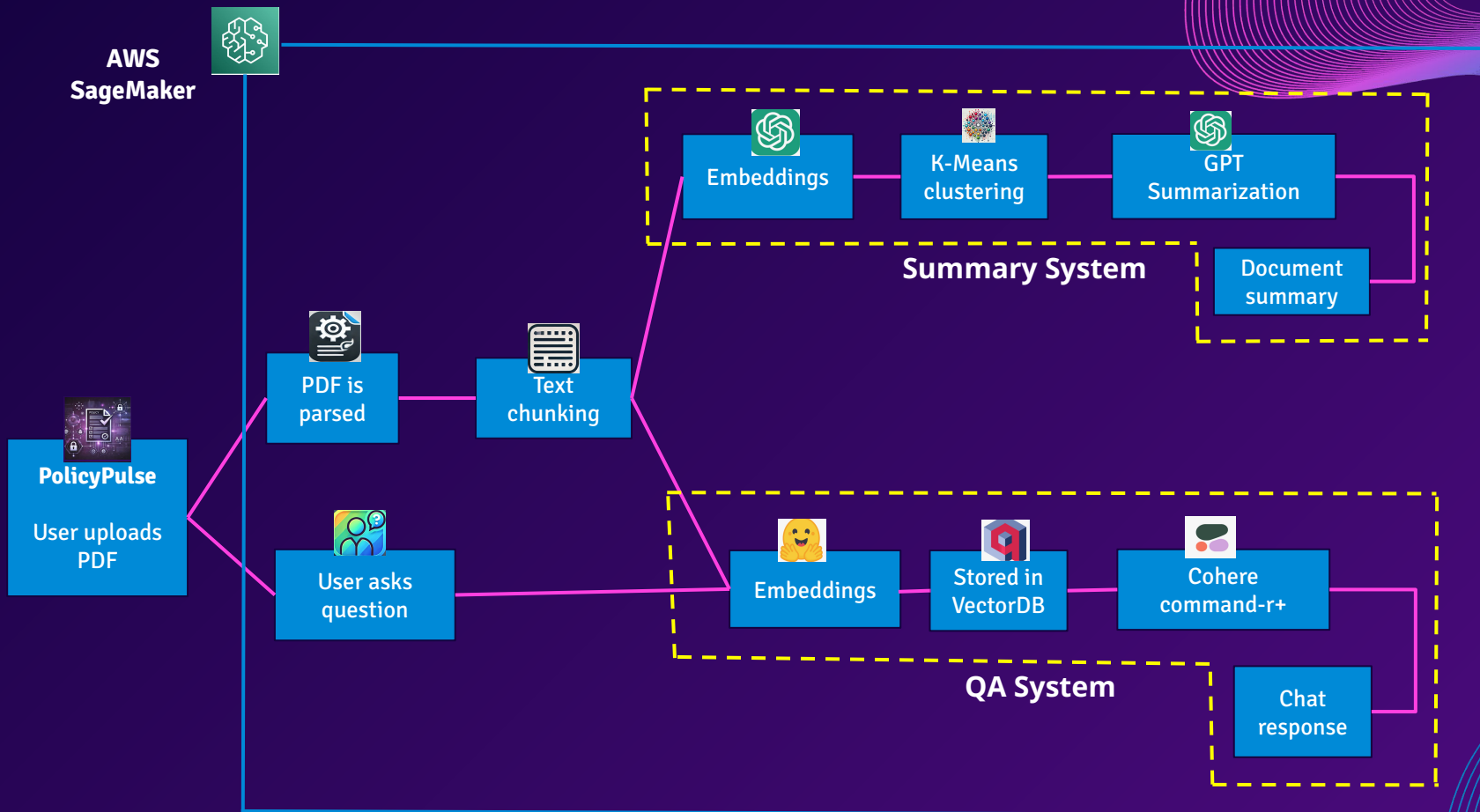
GPT-4o

Falcon

Legalbert

Deberta

# Architecture



# Ground Truth Generation

## Input Sources

10 Fintech  
Policies

GDPR Text

## Processing

### Question Types

What • When • Where  
Why • How

### Difficulty Levels

Easy • Medium • Hard

## Generated Q&A

(SME Verified)

### Fintech Questions (90)

9 per policy • Equal difficulty • Mixed types

### GDPR Questions (40)

20 Medium / 20 Hard • With recital references

bitpay

BILT

GEMINI

Klarna

PLAID

Revolut

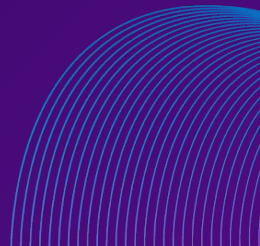
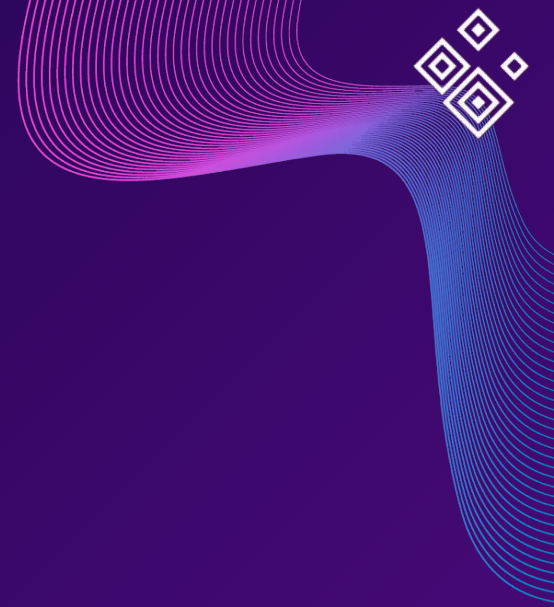
Remitly

stripe

MoonPay

CIRCLE

# Exploratory Data Analysis





# International policies are 5x longer than average





# GDPR questions fall only under Medium and Hard

Company	Easy	Hard	Medium
Bilt	2	3	3
BitPay	3	3	3
Circle	2	2	3
GDPR	0	20	20
Gemini	3	3	3
Klarna	2	3	3
Moonpay	2	4	2
Plaid	3	2	2
Remitly	3	3	2
Revolut	3	3	3
Stripe	3	3	3

# Easy to Hard Compliance Queries



## Easy

“What types of personal data does Bilt collect for rewards services”?



## Medium

“What rights do Bilt cardholders have regarding their personal data”?



## Hard

“How should organizations handle automated decision-making restrictions”?



# Summary Model Evaluation

Model	Version	SME scoring	ROUGE (Recall-Oriented Understudy for Gisting Evaluation )	Comments
Mistral	7B-Instruct-v0.2	Good	precision=0.2847222222222222, recall=0.1659919028340081, fmeasure=0.20971867007672634	The model's performance is hindered by hallucination and gives lengthier issues.
BART	-	Bad	precision=0.1958041958041958, recall=0.11382113821138211, fmeasure=0.14395886889460155	Least effective model. Excessively long summaries.
GPT	4o	Best	precision=0.5625, recall=0.32793522267206476, fmeasure=0.4143222506393862	Best overall performance, not too slow and gives concise summaries

# Summary Evaluation Rubric

## Score

Use Rouge score to compare against multiple models

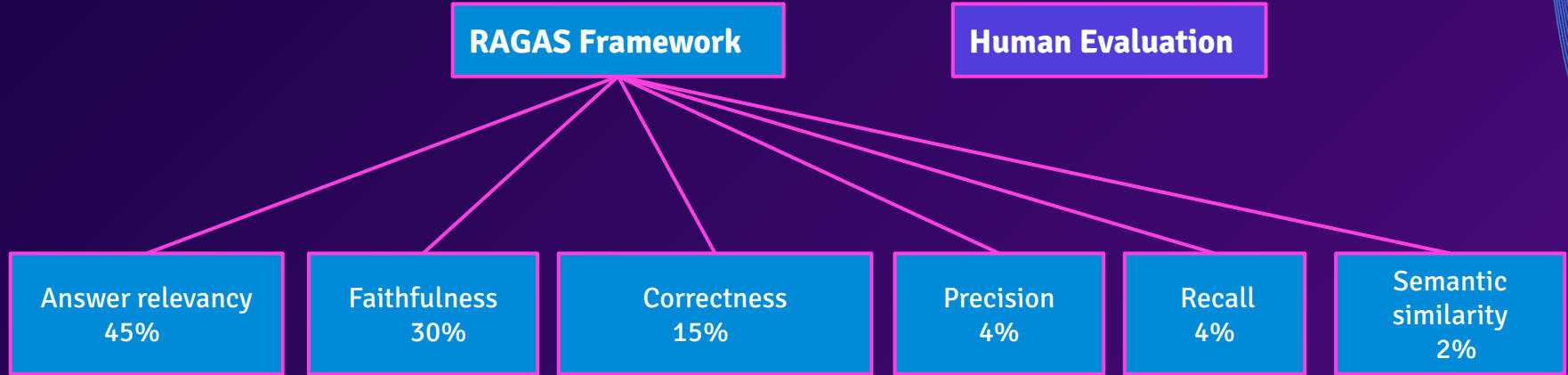
## Advantage

Summaries are tuned for our specific use cases

## SME

Help to get ground truth from User

# QA Evaluation Rubric



# QA Model Evaluation

Model	Version	Weighted Average (%)	Relevancy	Faithfulness	Correctness	Context precision	Context recall	Semantic similarity	Comments
Mistral	7B-Instruct-v0.2	68%	83%	62%	42%	75%	24%	90%	High accuracy but latency issues
Llama	Llama 2-7b	66%	75%	72%	45%	70%	21%	88%	Excels at medium complexity queries but slowest response times
Cohere	command	70%	78%	76%	48%	72%	24%	90%	Balanced performance with good speed
Cohere	command-r-plus	72%	78%	77%	52%	75%	28%	92%	Top performer across all metrics - best balance of accuracy and speed

# How is PolicyPulse Different?

Task	Other Chatbots	PolicyPulse
QA	Policy-based analysis: Answers questions only on policy text.	GDPR-based analysis: answers policy questions by connecting to GDPR articles and answers GDPR-specific questions by referencing recitals
Summary	Incomplete use of document, missing information, no ground truth so uses generic summaries	Harnesses map-reduce to handle long policy documents and crafts concise summaries
GDPR Checklist	Gives superficial checklist of GDPR article references which may not be accurate	Provides policy categories with mappings to GDPR articles for compliance check



# GDPR CHECKLIST IN ACTION

# Policy Category Analysis

Loaded policy from company:bitpay

Unique Policy Categories Found in the Policy :

- Data Retention
- Data Security
- International and Specific Audiences
- Other
- User Access, Edit and Deletion
- User Choice/Control

Policy Category Explanations:

- Data Security: How user information is protected.
- Other: Additional sub-labels for introductory or general text, contact information, and practices not covered by the other categories.
- Data Retention: How long user information is stored.
- User Choice/Control: Choices and control options available to users.
- International and Specific Audiences: Practices that pertain only to a specific group of users (e.g., children, Europeans, or California residents).
- User Access, Edit and Deletion: If and how users may access, edit, or delete their information.

# GDPR Statistics

## GDPR Article Analysis:

---

### 1. Policy Categories and Their Articles in the Uploaded Policy:

---

#### International and Specific Audiences:

Articles: Article 3, Article 27, Article 85, Article 87, Article 88, Article 91

#### Other:

Articles: Article 4, Article 31, Article 51, Article 52, Article 53, Article 54, Article 55, Article 56,

#### Data Retention:

Articles: Article 5

#### User Choice/Control:

Articles: Article 7, Article 8, Article 18, Article 21, Article 22, Article 23, Article 77, Article 78,

#### User Access, Edit and Deletion:

Articles: Article 15, Article 16, Article 17, Article 20

#### Data Security:

Articles: Article 25, Article 32, Article 33, Article 34, Article 35, Article 36, Article 37, Article 38

### 2. Missing GDPR Articles:

---

#### First Party Collection/Use:

Articles: Article 1, Article 2, Article 6, Article 9, Article 10, Article 11, Article 12, Article 13, Article 14, Article 2

#### Third Party Sharing/Collection:

Articles: Article 19, Article 28, Article 44, Article 45, Article 46, Article 47, Article 48, Article 49, Article 50





# Key Learnings

## Ground Truth Q&A

Iterative Process w/ SME

## Model Improvement

Incremental Changes

## Model Evaluation

Command R-Plus Performance

## Inference & Integration

Debugging & Infrastructure

# Future work

Dataset  
refinement for  
ground truth

Compliance  
matching  
against GDPR

Checklist  
Integration/  
Improvement

Knowledge  
graphs  
mapping

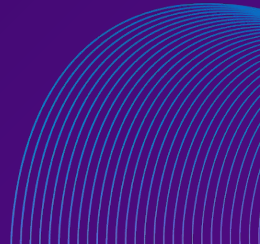
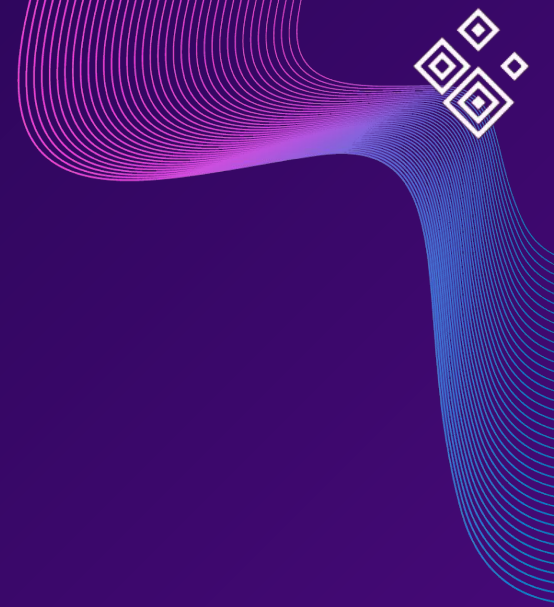


# Mission Statement

*“Unlock seamless data privacy compliance with our cutting-edge AI solutions, empowering fintechs to identify risks early, safeguard user data, and foster unwavering trust in the digital age.”*

# Acknowledgements

- Kevin Kuc: MIDS student
- Jared Maslin: MIDS 231 Privacy compliance faculty
- Mark Butler: MIDS 266/267 NLP and Gen AI faculty
- Timothy Cradle: Fintech compliance professional
- Tim Tung: MIDS student
- Samantha: Prior MIDS student
- Nayan: MIDS student
- Saurabh: MIDS student
- Billy Fong: MIDS student
- Joyce Shen: MIDS 210 faculty
- Korin Reid: MIDS 210 faculty



# References

- Privacy Pulse project title image: AI generation GPT
- Ground truth Q&A generation with ChatGPT and Legal GPT
- Slides template from [Slidesgo](#)
- PrivacyPulse logo (AI tool)
- RegTech visual inspired by [Fortune Business Insights](#)
- GDPR fines