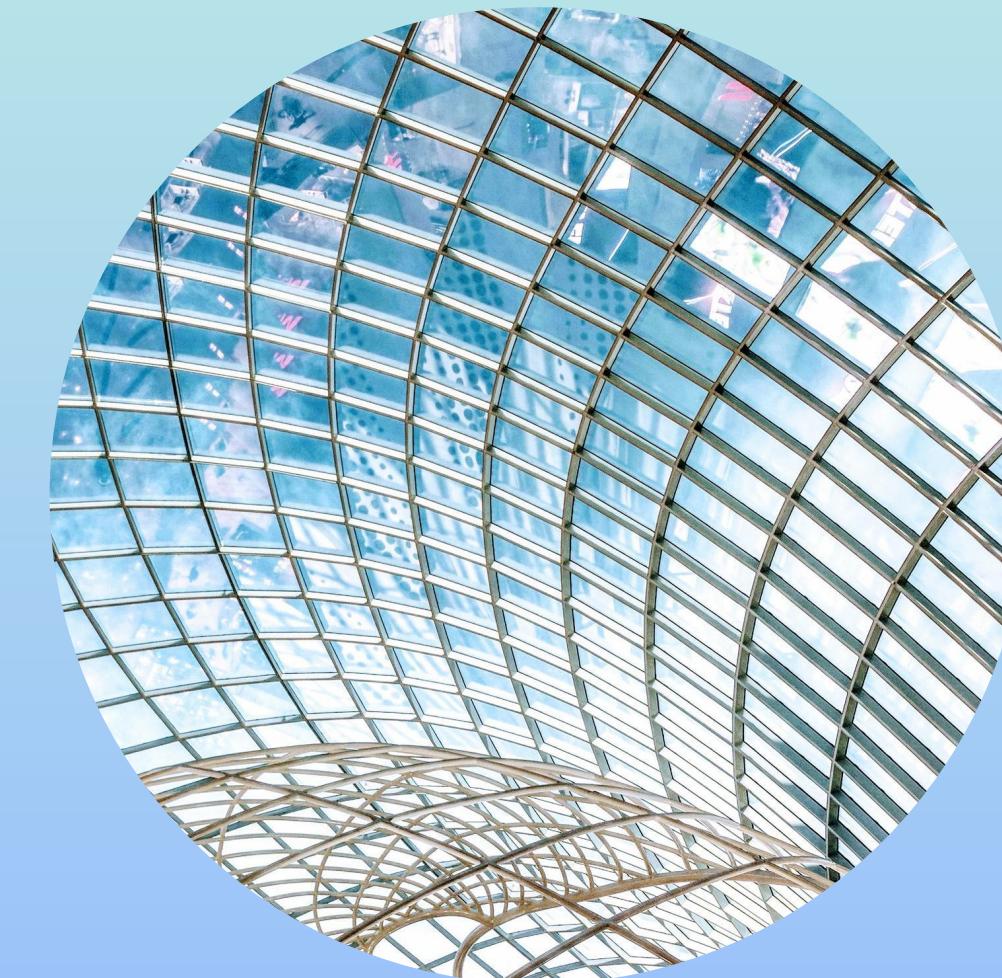




DL

GROUP 10

Agustin, Birk, Karol



Team Members



Agustin



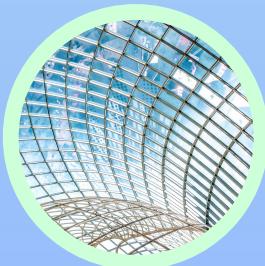
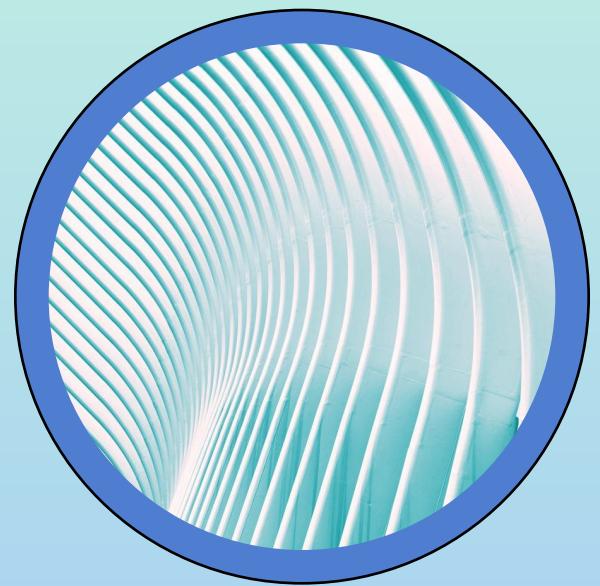
Birk



Karol

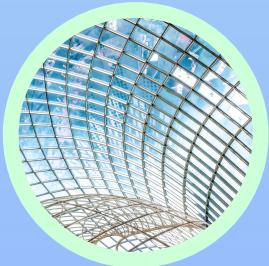
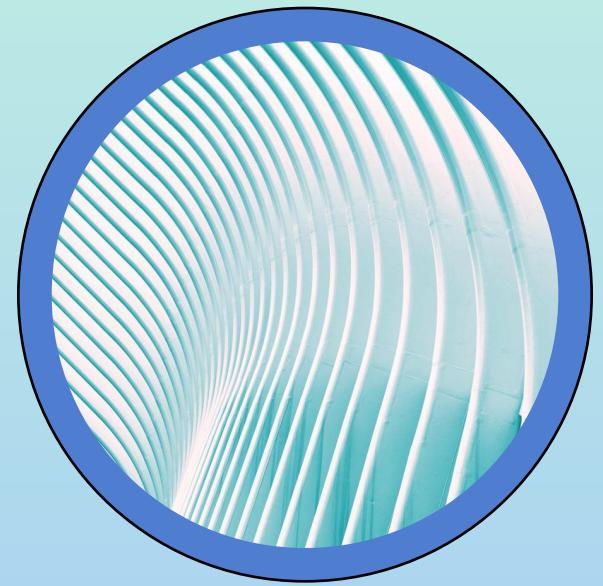
Table of Contents

- **System Context**
- **Model Overview**
- **Model Card**
- **Dataset Card**



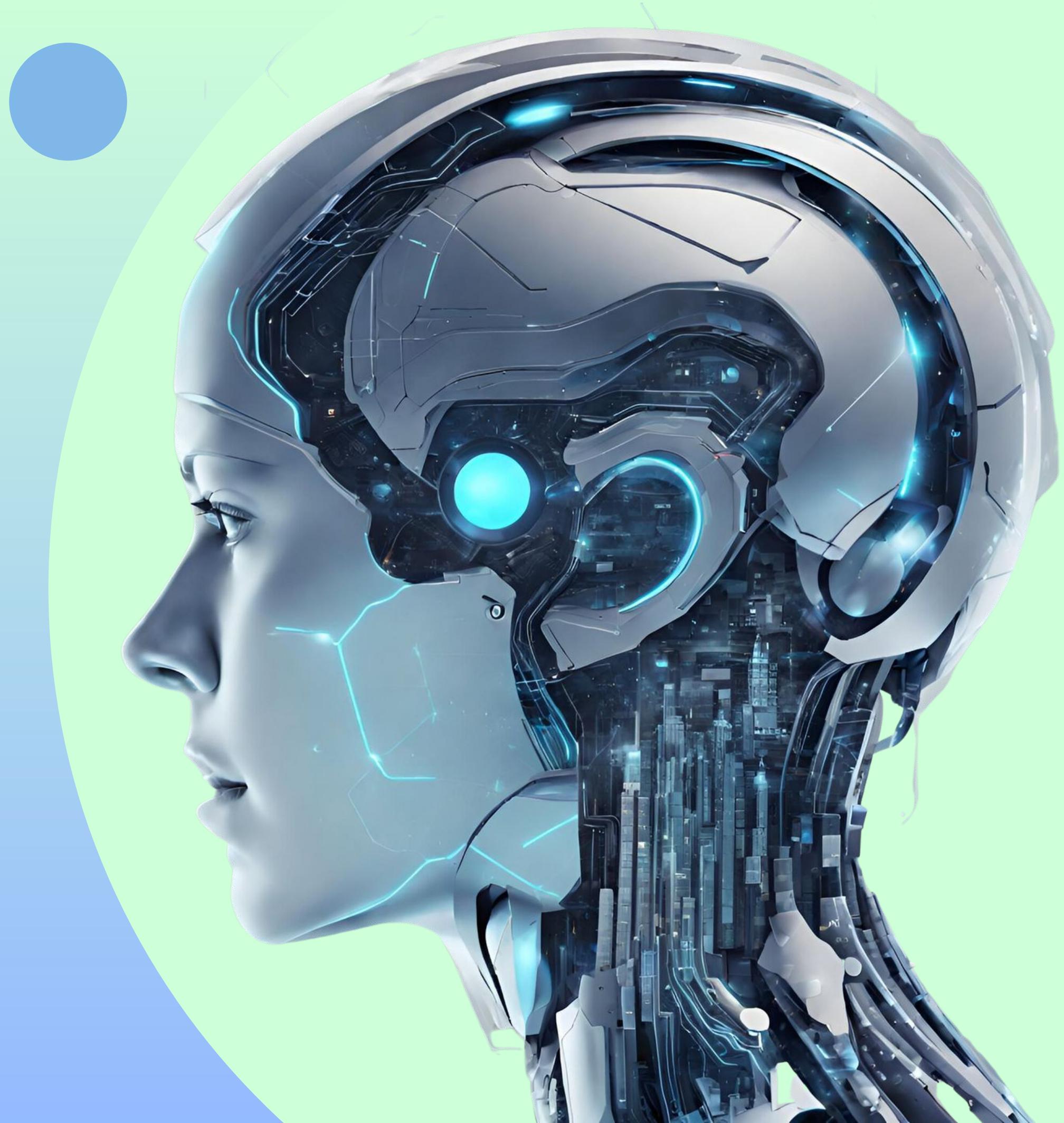
System Architecture Summary

- **Web application** for uploading or pasting articles
- **Backend API** handles classification requests
- Preprocessing module **cleans** & **vectorizes** text (**TF-IDF**)
- **Dense Neural Network** model produces probability score
- Results and metrics stored for **monitoring** & **evaluation**



How the System Works

1. User submits article (**title + text**)
2. Text **cleaned & converted** into features
3. Model predicts **Fake/True** + **confidence score**
4. Interface shows **results** and **key words**
5. Logs saved for **traceability** and **review**

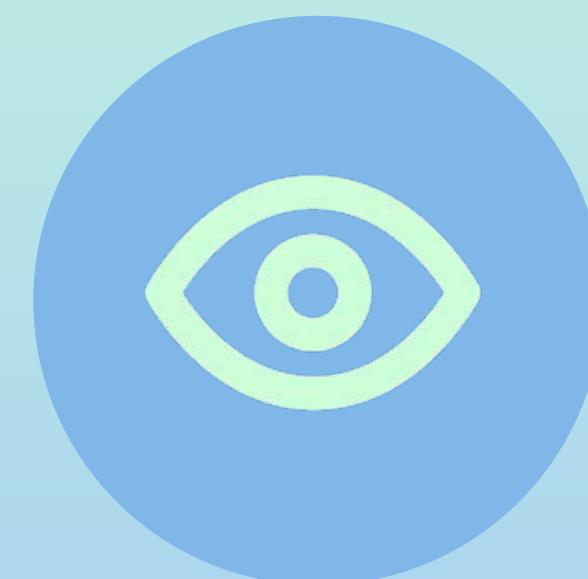


Core System Functions



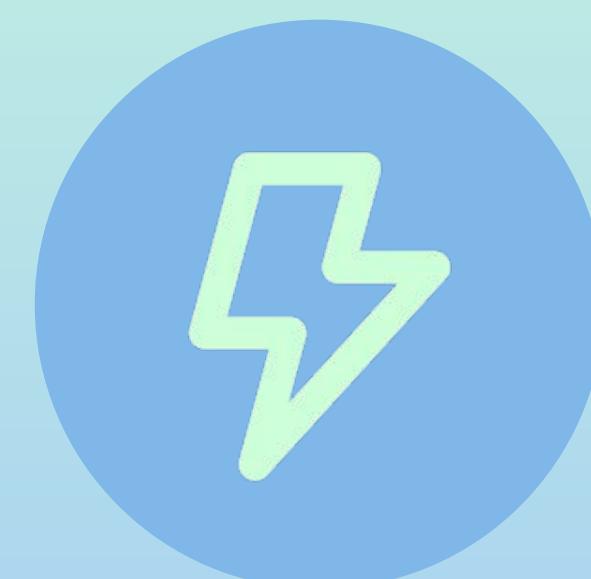
**Accept article input
via web page or API**

Clean & standardize
text before inference



**Produce Fake/True label
+ confidence (0–1)**

Categorize as Likely
Fake / Uncertain /
Likely True

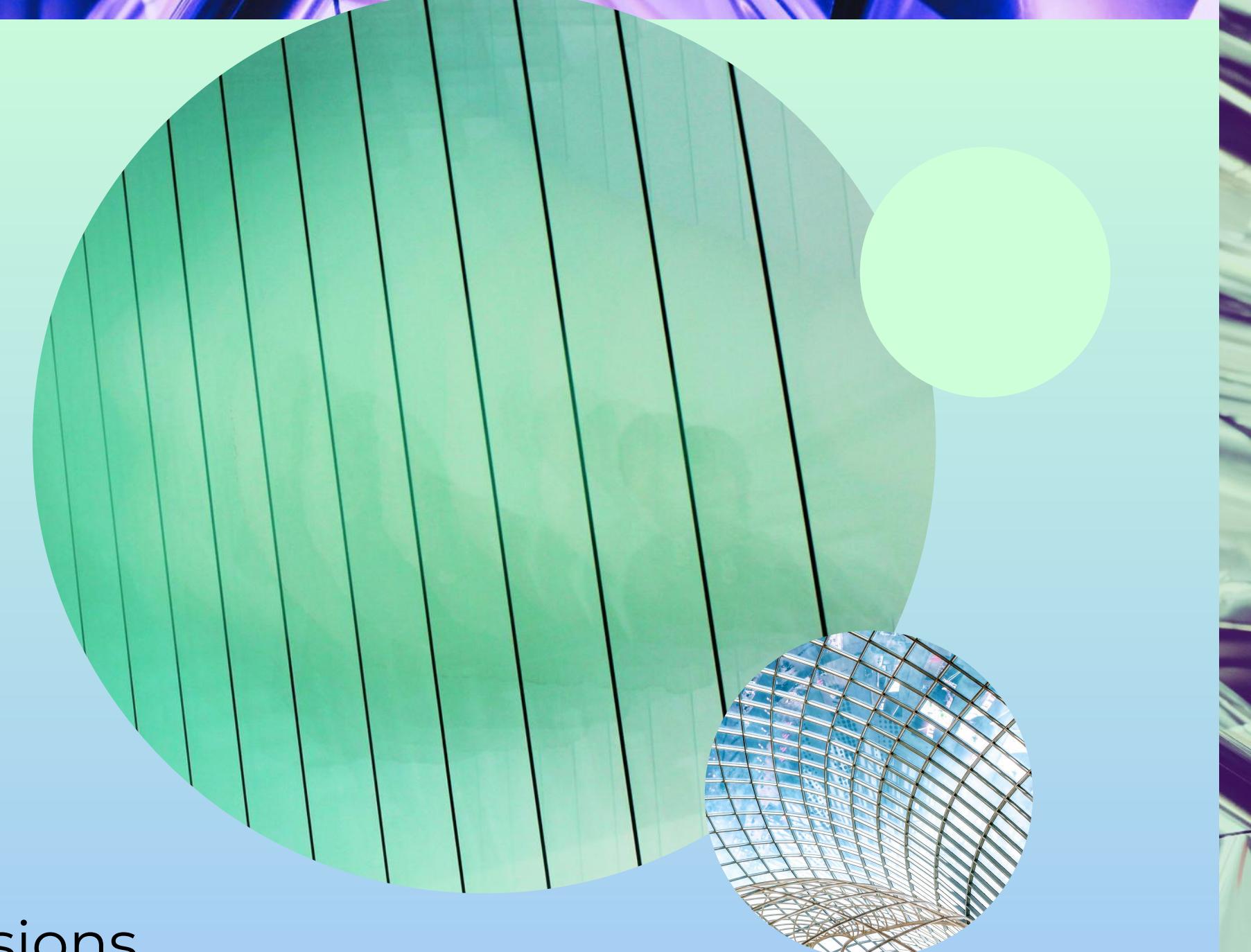


**Validate API
inputs**

Log anonymized
results

Performance & Quality Constraints

- **Inference $\leq 3 \text{ s}$** per article (CPU)
- **Web UI** shows results $\leq 2 \text{ s}$ after prediction
- **~95% uptime** during testing sessions
- **Secure HTTPS** communication + auth **tokens**
- Modular design → easy **maintenance** & future **scaling**



Performance Goals

- **F1-score**
≥ 0.85 on test set
- **Deterministic training**
with fixed seeds
- **Each prediction**
≤ 3 s on laptop CPU
- **Review ≥ 10**
misclassified samples
for analysis



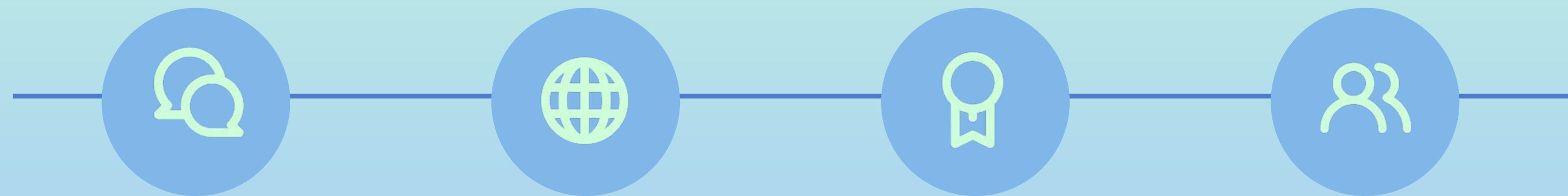
Challenges & Solutions

- **Bias:** political tone → false fake detections → evaluate by subject
- **Imbalance/leakage:** uneven classes → use stratified splits + regularization
- **Overfitting:** sparse features → apply dropout & early stopping

Responsible AI Principles

- **Educational tool** – not a fact-checking engine
- Shows **confidence & influential words** for transparency
- Processes data **locally** (no PII stored)
- Follows **ethical AI** and **human-oversight** guidelines

Future Development (M2-M3)



Replace Dense NN

with DistilBERT for
context awareness

Expand API

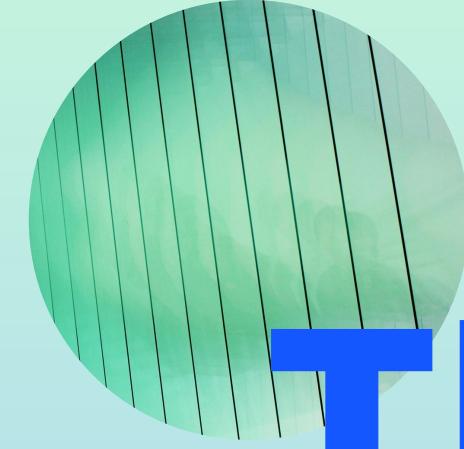
Deploy full web
application

Dashboard

Implement continuous
evaluation dashboard

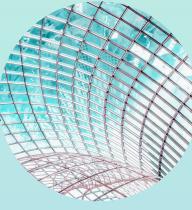
Integration

Integrate fact-checking
and source credibility
modules



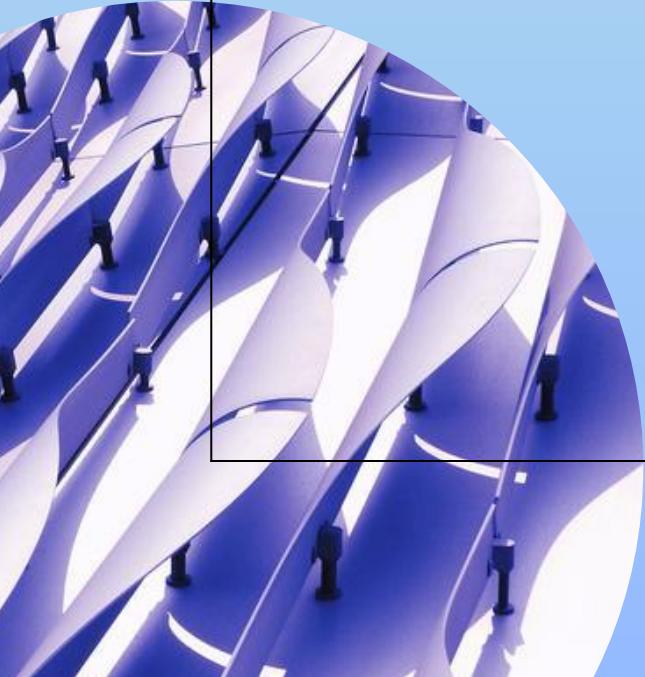
The chosen model and the reasons behind the choice

Model Overview

			
Goal	Input	Output	Representation
Classify news articles as real or fake based on their text.	Concatenated title + body of each article.	Probability of the article being fake → binary classification.	Text → TF-IDF vectors. Captures important words efficiently.

Model: Feed-Forward Dense Neural Network

- Layers: Fully connected (Dense) + ReLU activations
- Dropout for regularization
- Sigmoid output for fake/real probability
- Trained with Binary Cross-Entropy + Adam optimizer



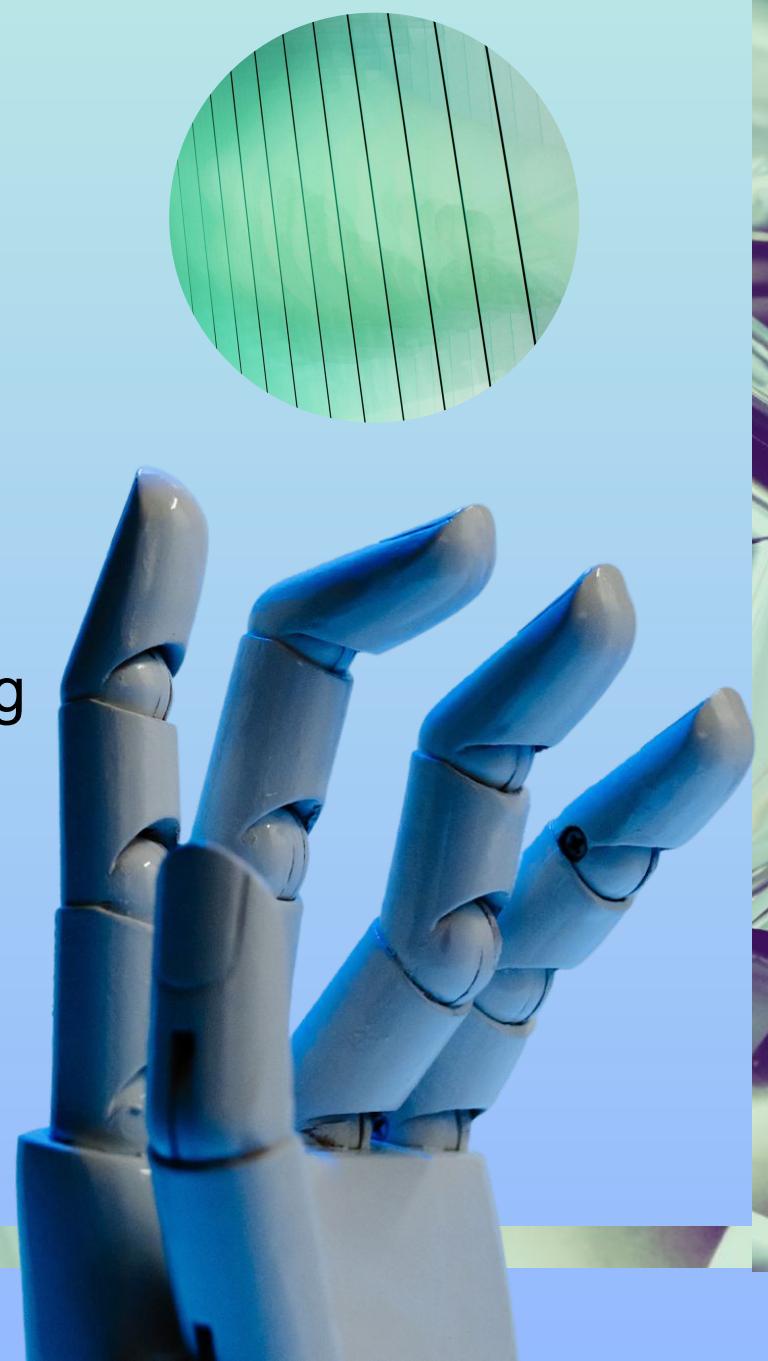
Models considered

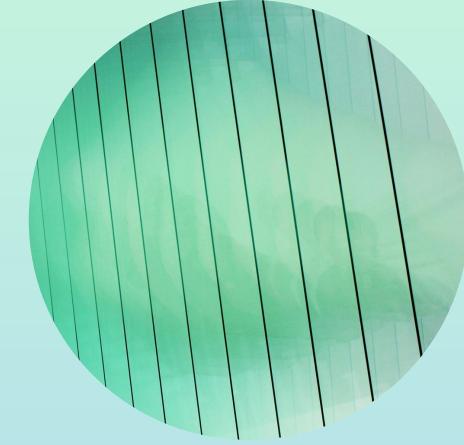
- **RNNs (LSTM / GRU)**: capture sequence and word order
→ *Accurate but slow and resource-heavy.*
- **CNNs**: detect local word or phrase patterns
→ *Efficient but limited for long-range context.*
- **Transformers (BERT / DistilBERT)**: best contextual understanding
→ *High accuracy, but require more data and compute power.*



why This Model & What's Next

- **Chosen model:** Feed-forward Dense Neural Network
 - Simple, interpretable, and efficient for TF-IDF text data.
- **Why it fits:**
 - Balanced trade-off between simplicity and expressive power
 - Easy to reproduce and compare in future experiments
- **Next steps (Milestones 2–3):**
 - Explore **Transformer-based models (DistilBERT)** for contextual understanding
 - Compare performance and robustness
 - Deploy best model in final software system

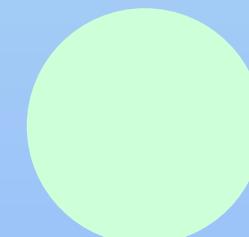
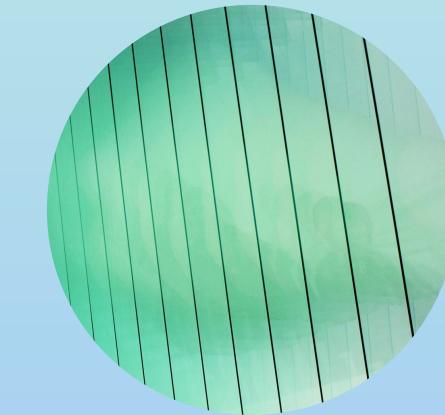
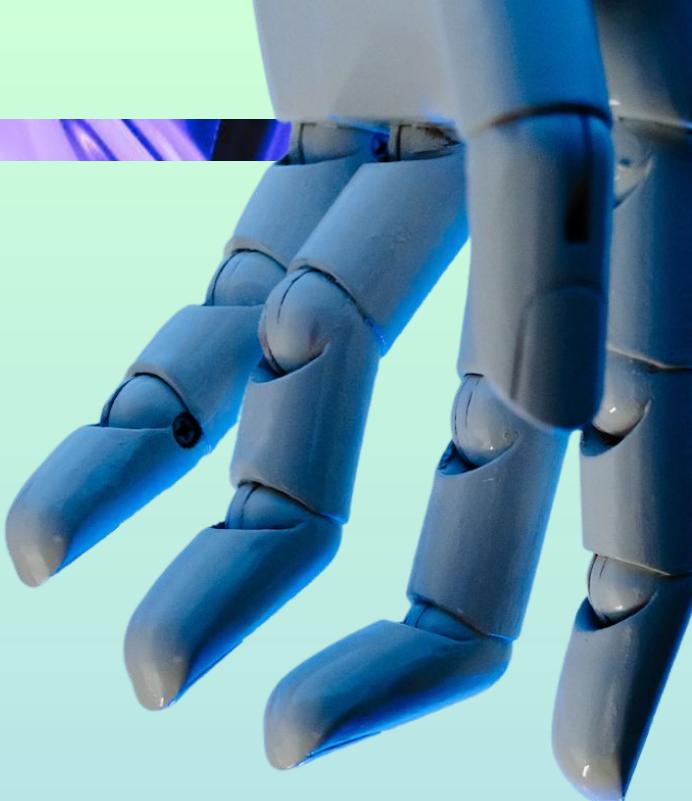




Model & Data Cards

Transparency in DL: Model & Data Cards

- Model and Data Cards help us explain **how** and **why** a DL system works.
→ Both are documentation tools for ML transparency.
- **Model Card:** explains *how* the model works, its purpose, and ethical limits.
- **Data Card:** describes *what* data was used, its source, and possible biases.
- They help us insure DL systems are **responsible, fair, and easier to understand.**



Model Card - Summary

- **The goal** of the model is to find out if a news article looks *real* or *fake* based on how it's written.
- **Intended Use:**
 - Educational tool to study misinformation (not a fact-checker).
 - Designed to aid users increase awareness about misinformation.
- **Planned results:** We want the model to reach around **85% accuracy (F1-score)**.
- **Factors:** The topic of the article, the writing style, and the fact that all the text is in English, may cause variation.
- **Ethics:** Must be reviewed by humans.
 - Could reflect dataset biases.

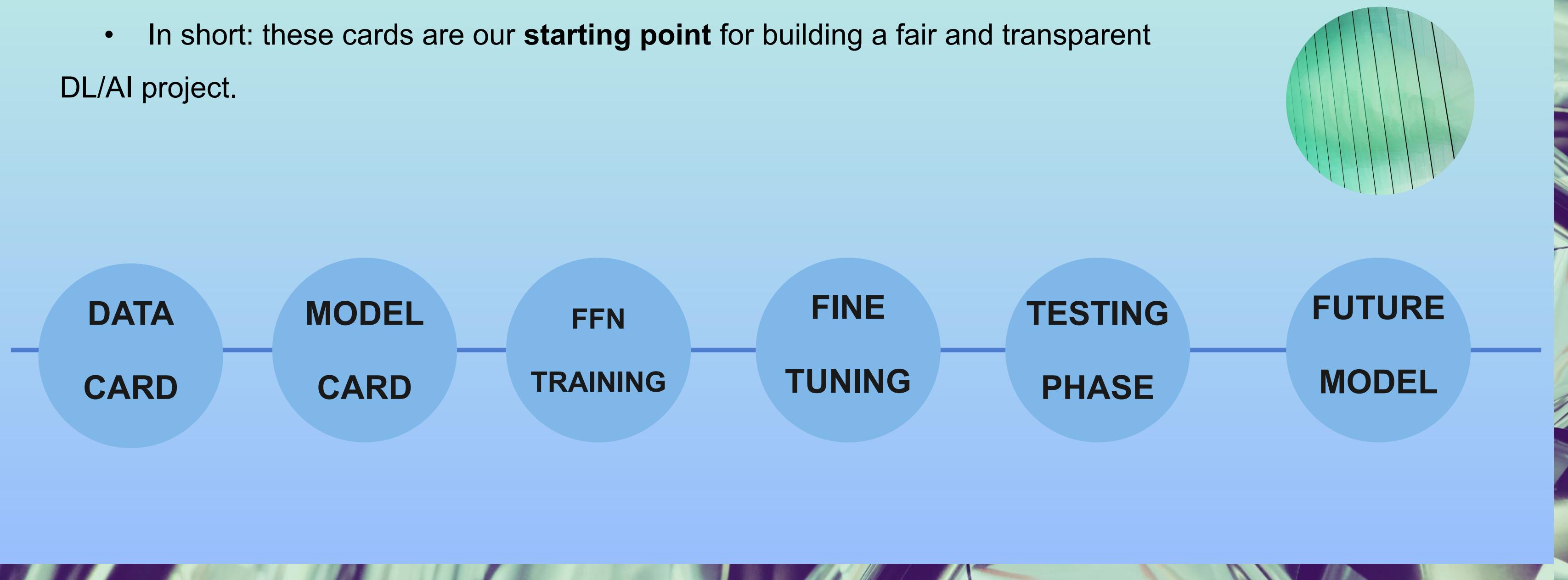


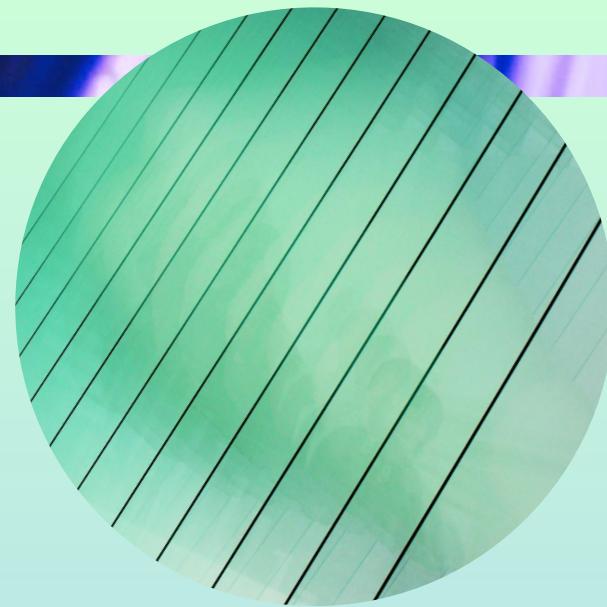
Dataset Card - Summary

- **Source:** Kaggle dataset by Clément Bisaillon (based on ISOT).
- **Composition:** ~44 000 English articles – columns (ID, title, text, subject, date, label).
 - There is good balance between real news (~45%) and fake news (~55%) in the dataset.
- **Purpose:** Train/test models that detect misinformation.
- **Splits:** 80 % train / 10 % validation / 10 % test.
- **Bias & Limitations:** Mostly political, English-only, writing style influences labels.
- **Ethics:** Public data; no personal information.

Key Takeaways & Next Steps

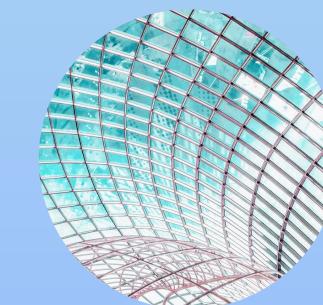
- These cards make our project **clearer** and **easier** to explain.
- They help us stay organized and aware of the model's limits.
- In short: these cards are our **starting point** for building a fair and transparent DL/AI project.

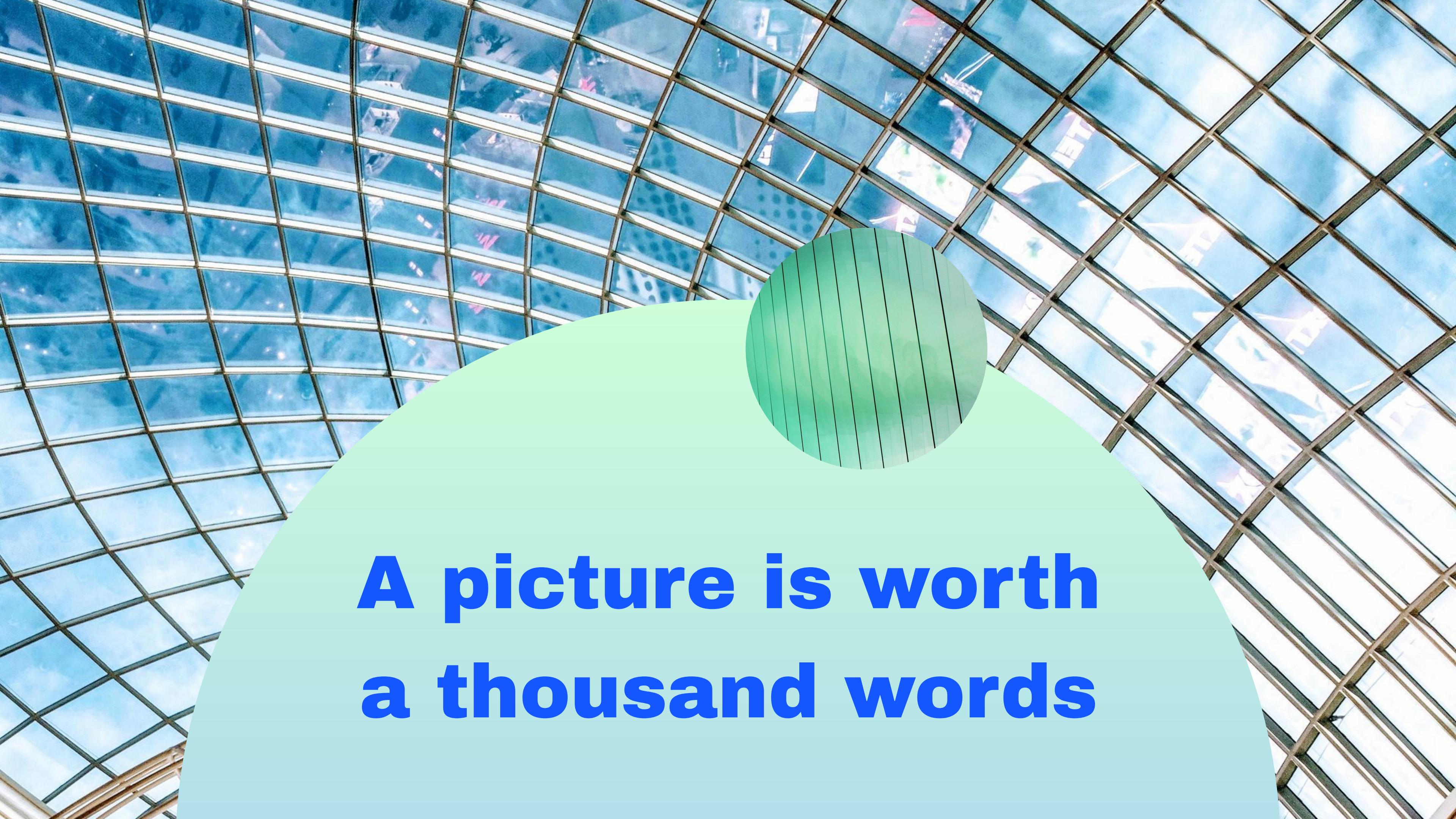




Thank you!

See you in Milestone 2!

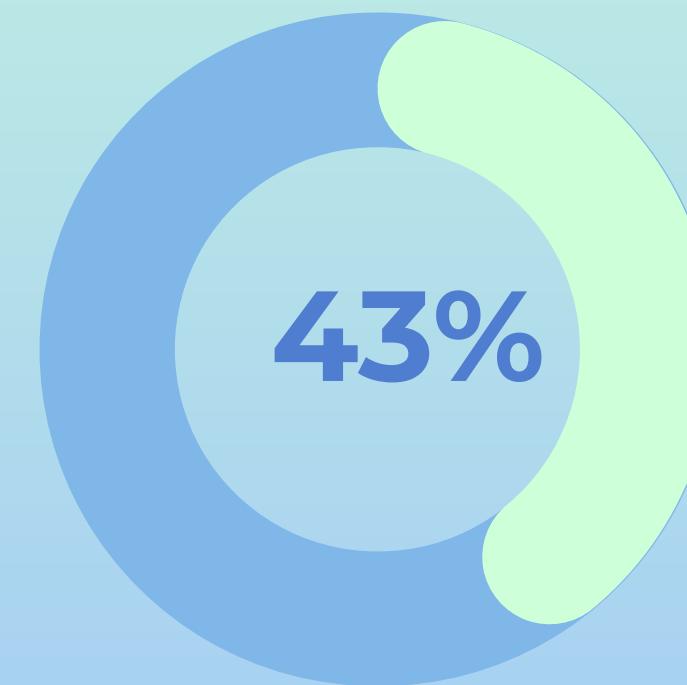




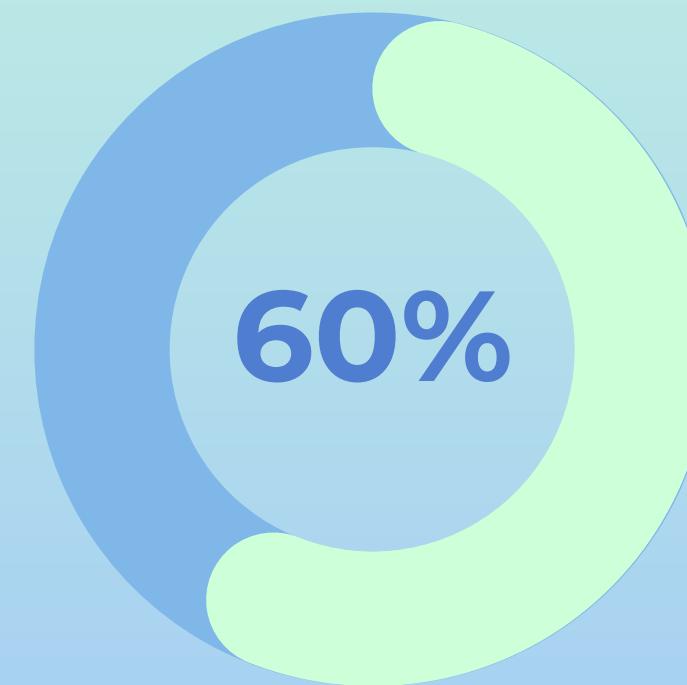
**A picture is worth
a thousand words**

Market Research

Elaborate on the featured statistic



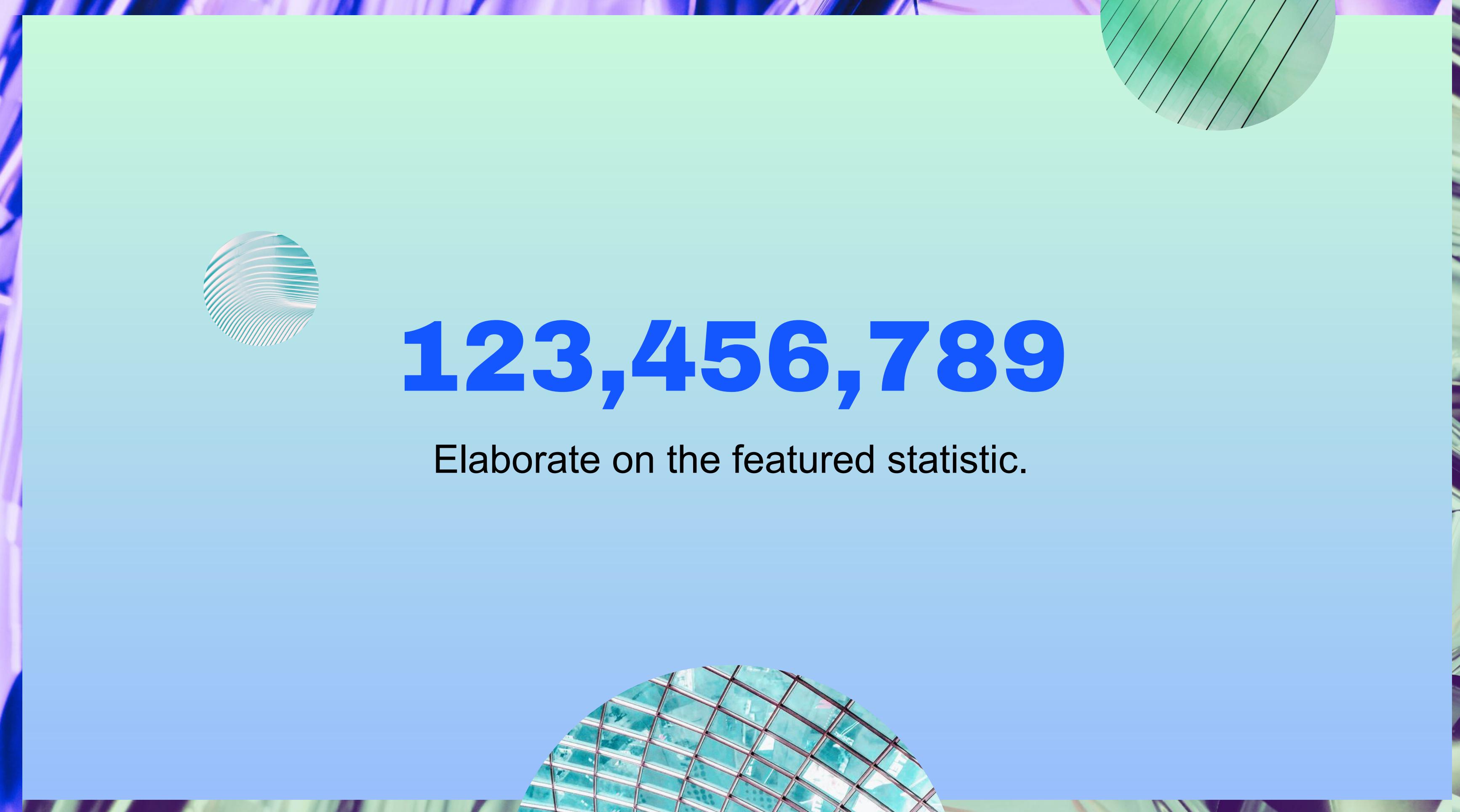
Elaborate on the featured statistic.



Elaborate on the featured statistic.



Elaborate on the featured statistic.



123,456,789

Elaborate on the featured statistic.

Resource Page

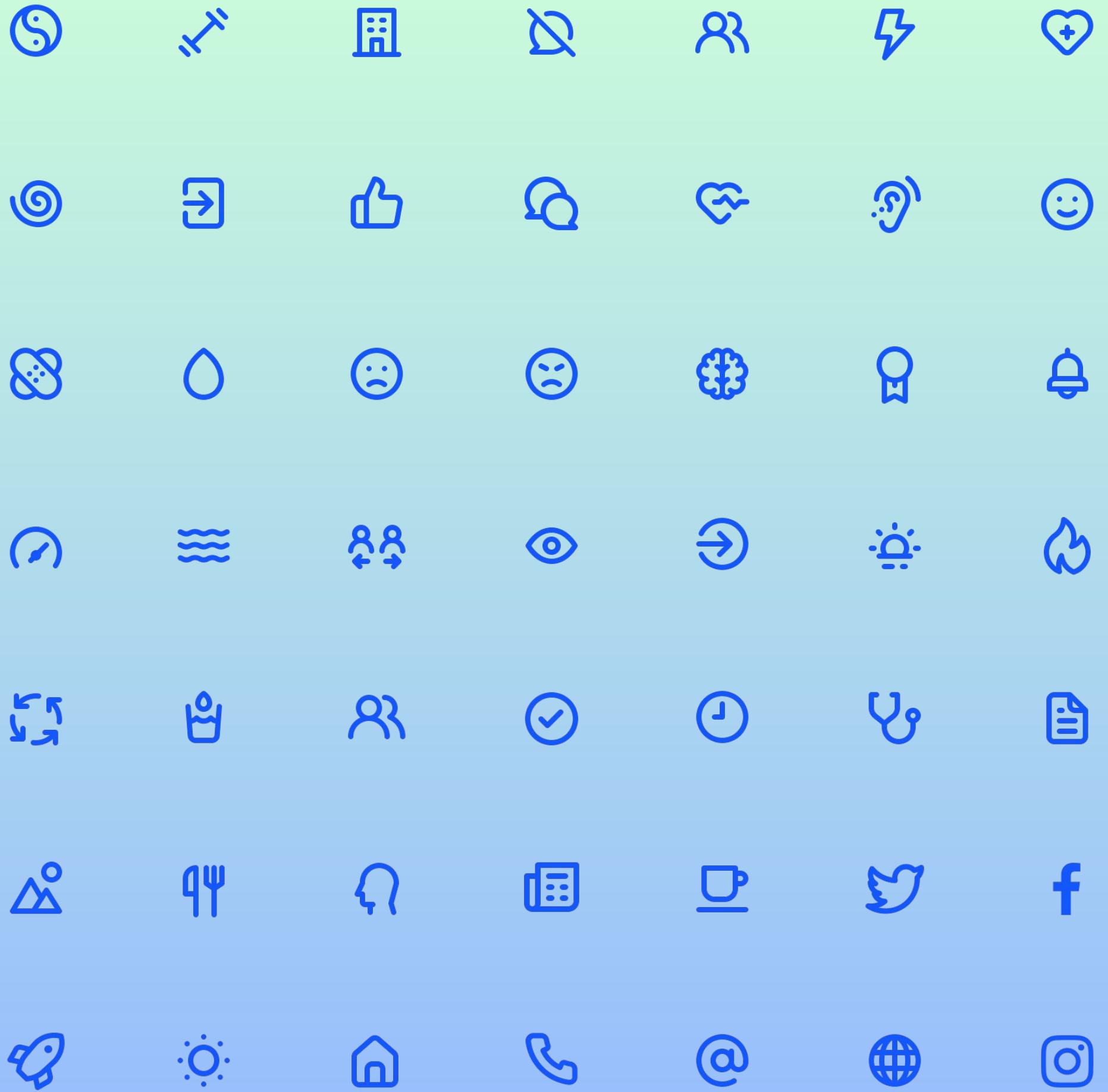
Use these design resources in your Canva Presentation.

This presentation template uses the following free fonts:

- Titles: Archivo Black
- Headers: Montserrat
- Body Copy: Open Sauce

You can find these fonts online too.
Happy designing!

Don't forget to delete this page before presenting.



Credits



This presentation template is free for everyone
to use thanks to the following:

SlidesCarnival for the presentation template
Pexels for the photos

Happy designing!