

# Enhancing ESG Data Collection: LLM Approaches for Airline Emission KPIs



<https://azero2050.eu/>

## **Luis MARTÍN-DOMINGO**

Business School, Dublin City University, Dublin 9, Ireland  
Faculty of Aviation, Ozyegin University, Istanbul, Turkey.

## **Jaime B. FERNANDEZ**

Insight Research Ireland Centre for Data Analytics, Dublin City University, Dublin 9, Ireland

## **Marina EFTHYMIU**

Business School, Dublin City University, Dublin 9, Ireland

## **Muhammad Intizar ALI**

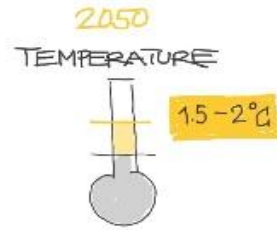
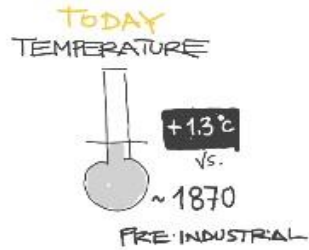
Insight Research Ireland Centre for Data Analytics, Dublin City University, Dublin 9, Ireland



**Funded by  
the European Union**

This project has received funding from the European Union's HORIZON-MSCA-2023-PF-01 research programme under the Marie Skłodowska-Curie (MSCA) grant agreement No 101151804

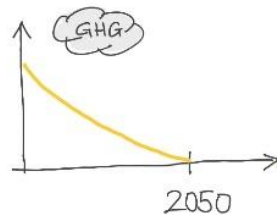
# Context



**Temperature increase** and suggested ceiling by science

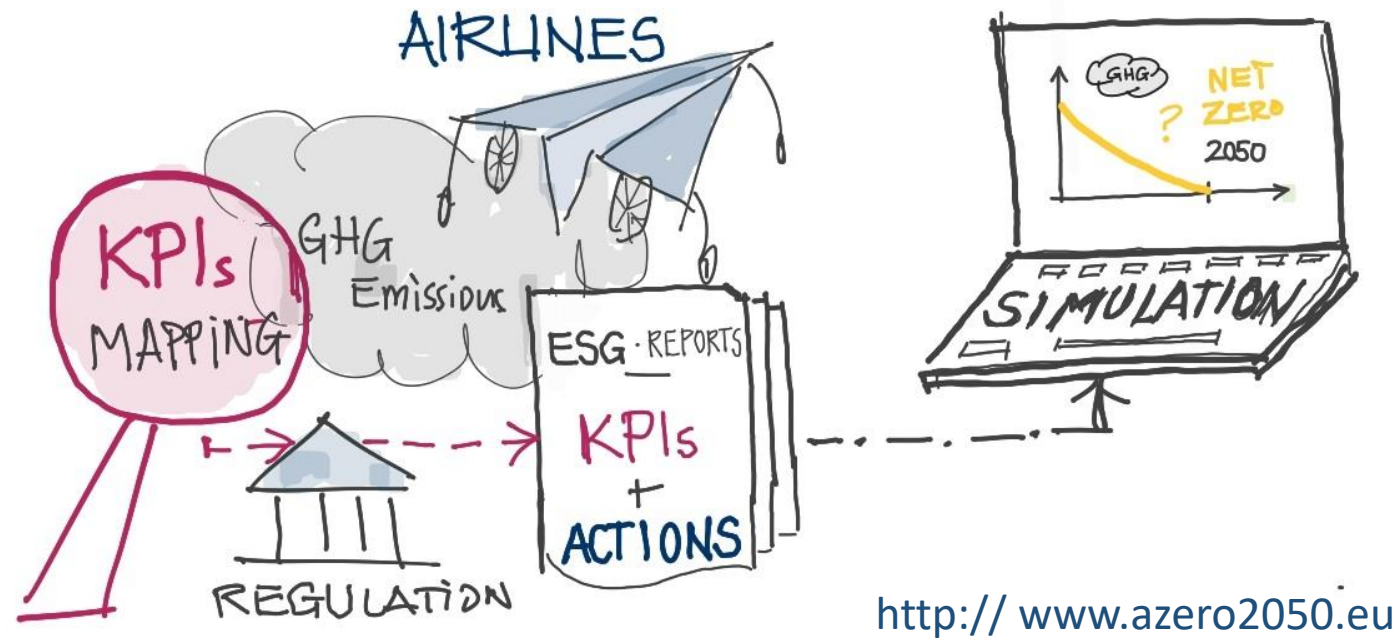


**Paris Agreement EU Green Deal** leading to **Net Zero 2050**



**Air Transport commitment** to Net Zero 2050, but researcher **concerns** (Klöwer et al., 2021)

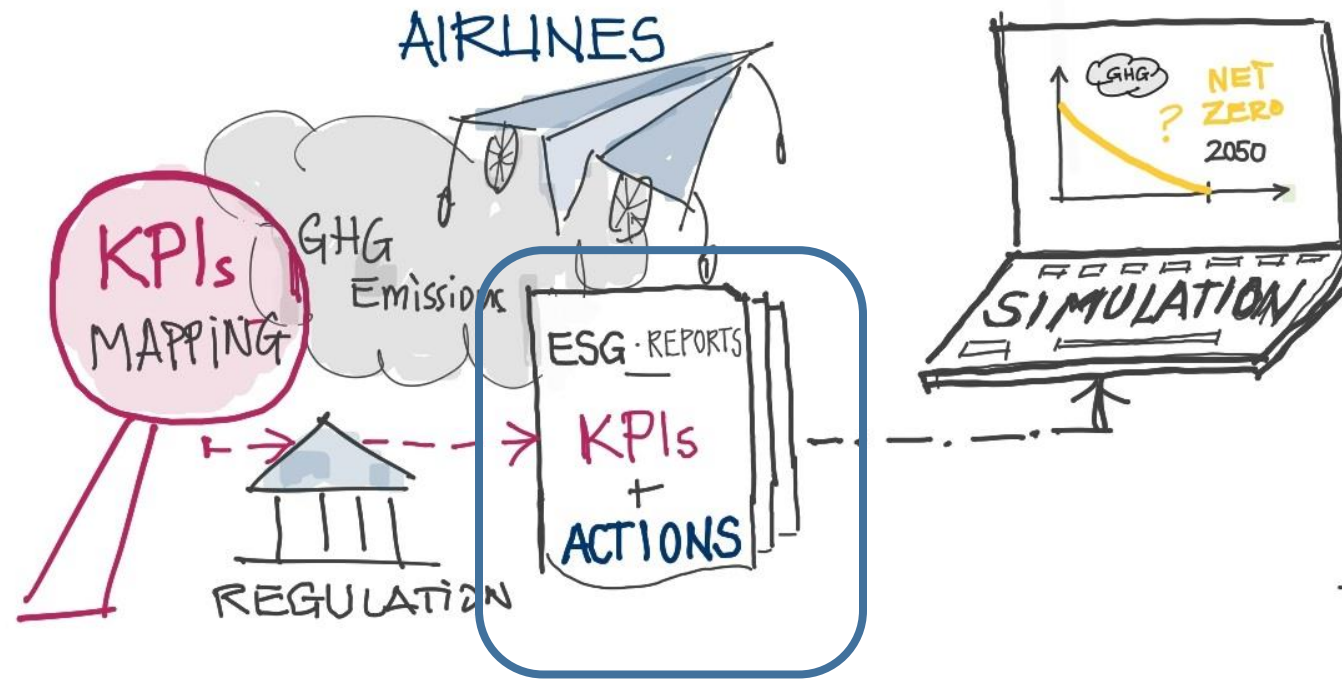
# Aviation Zero Emissions 2050 (AZERO) Project



Funded by  
the European Union

This project has received funding from the European Union's HORIZON-MSCA-2023-PF-01 research programme under the Marie Skłodowska-Curie (MSCA) grant agreement No 101151804

# Aviation Zero Emissions 2050 (AZERO) Project



Enhancing ESG Data Collection: LLM Approaches for Airline Emission KPIs



Funded by  
the European Union

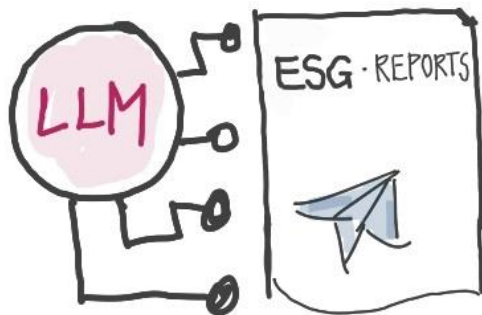
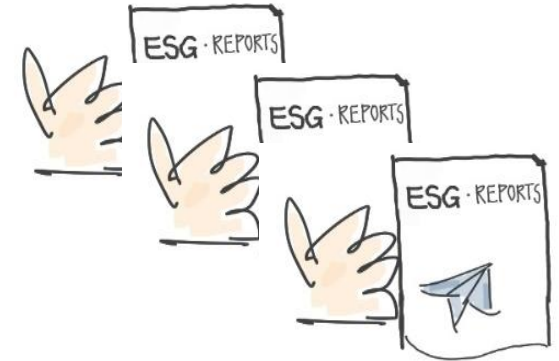
This project has received funding from the European Union's HORIZON-MSCA-2023-PF-01 research programme under the Marie Skłodowska-Curie (MSCA) grant agreement No 101151804

# Motivation



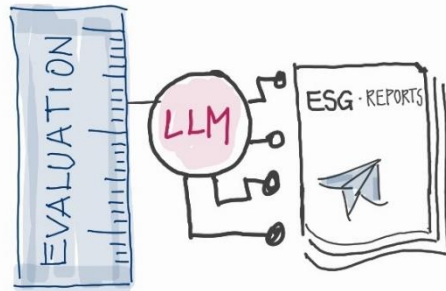
ESG ratings are **central to investment**, and can be used as a data source

**Manual extraction** from unstructured reports is **laborious** and **inconsistent**

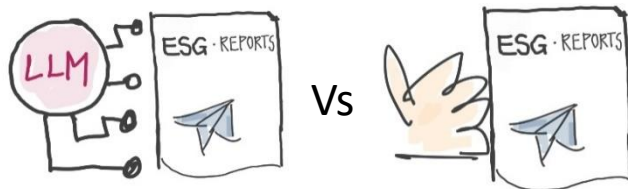


**Automated extraction** using Large Language Models (LLMs)

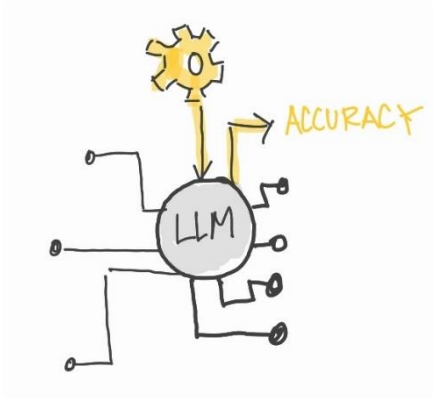
# Research Objectives



**Evaluate LLMs for automating emission KPI data extraction**



**Compare manual expert extraction with automated approaches**

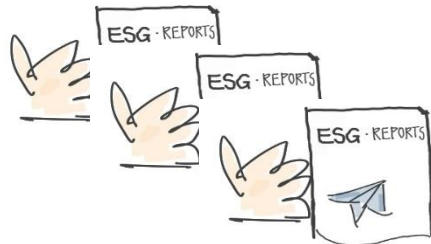


**Identify the factors influencing the accuracy of LLM extraction**

# Literature Review – Challenges in ESG Reporting



**Lack of standardization** and comparability in **airline sustainability reporting**

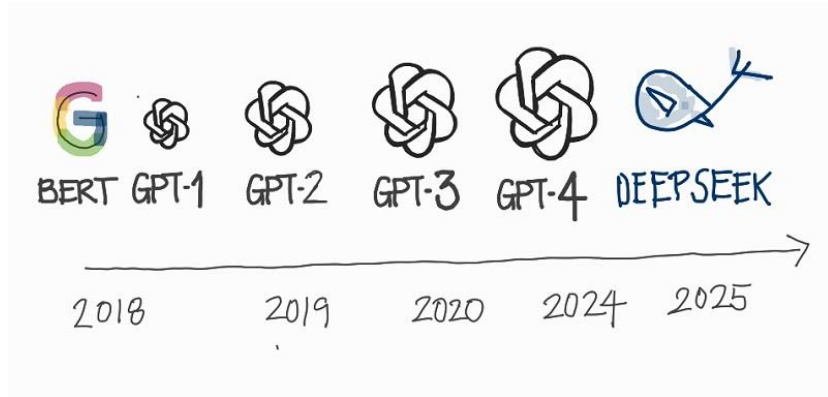


**Manual** content analysis is both: **resource-intensive** and **error-prone**



Emergence of **LLMs** as **promising** tools for **automating** ESG data extraction

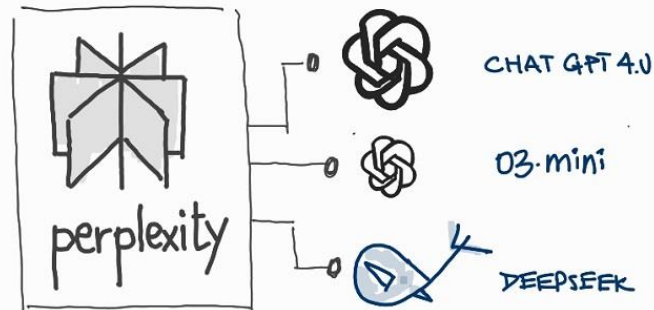
# LLMs in Context



**Brief history:** From BERT to GPT-4 and Deepseek R1



**Capabilities:** Context-aware extraction, handling unstructured data



**Accessibility:** Platforms like Perplexity enable practical application

# Methodology – Data Collection



**Identification** of the **16 European airline groups**

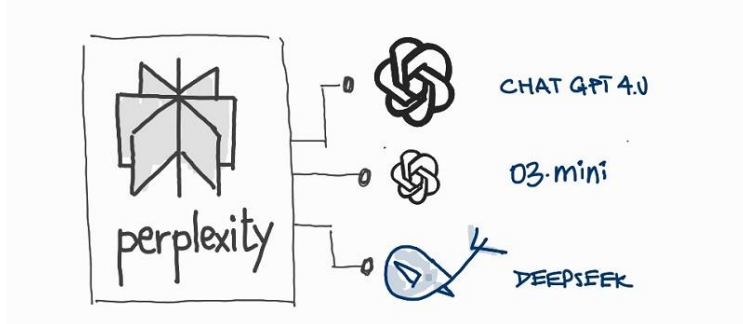


**Data sources:** Airline Annual and sustainability / ESG reports -- (PDF, HTML)

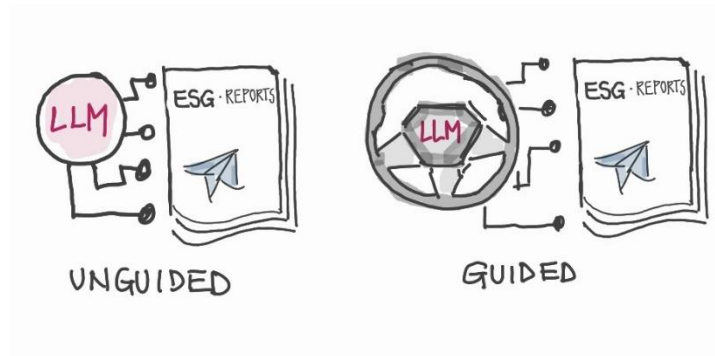


**Manual extraction:** **Two experts** reviewed and tabulated emission KPIs

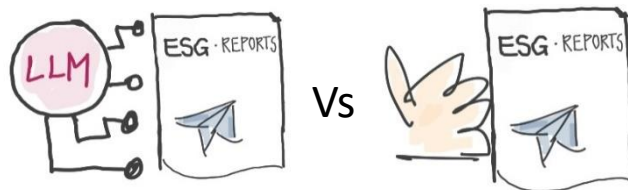
# Methodology – LLM Extraction Workflow



Three LLMs evaluated via Perplexity platform

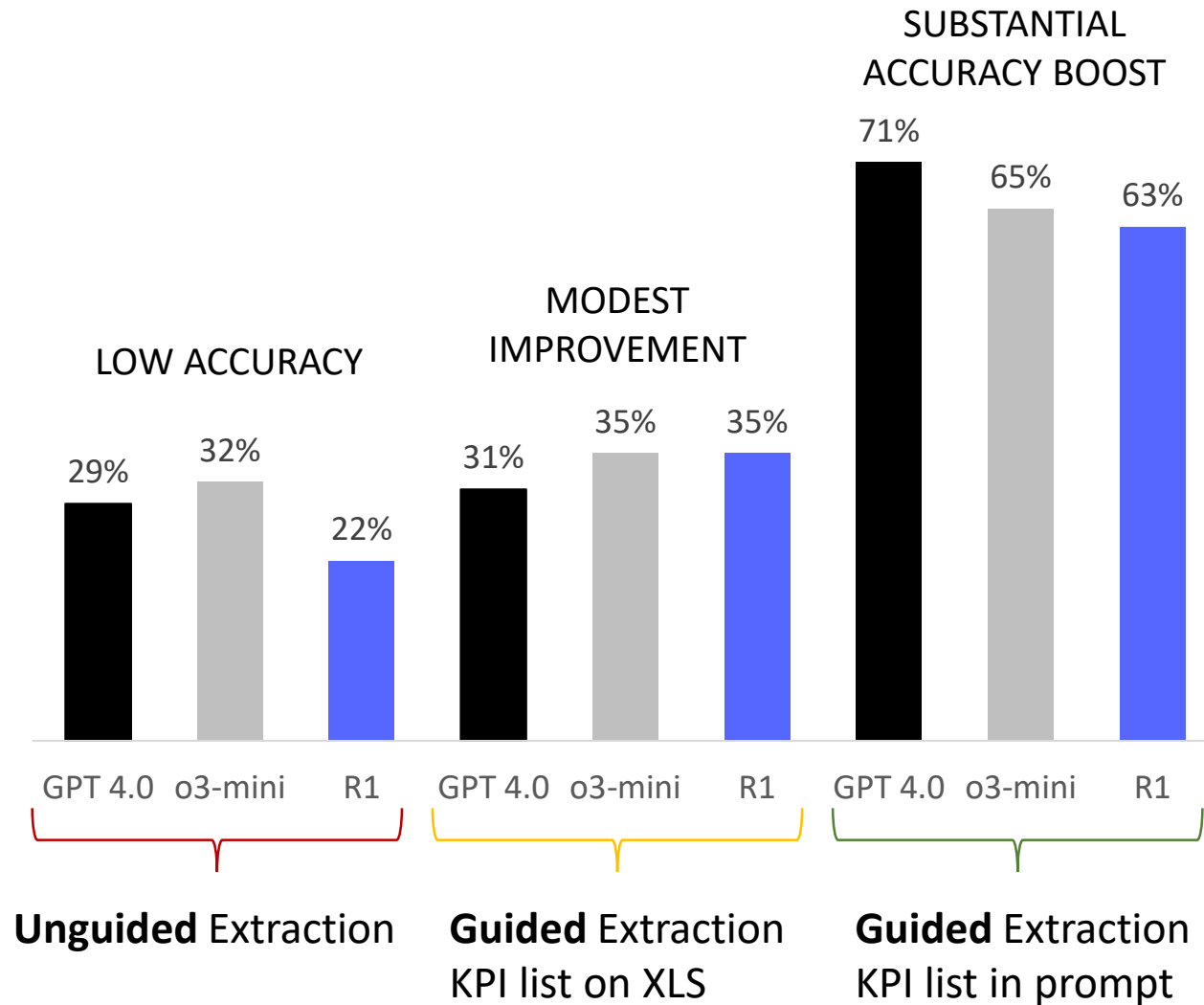


Extraction strategies: **Unguided** and **guided** prompts



Comparison of extracted **KPIs**: name, unit, value

# Results – Extraction Accuracy by Model & Prompt



Limited accuracy variation among the **LLMs** used

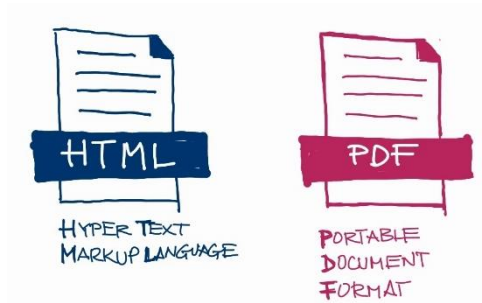
**Prompt specificity** is the **dominant factor** in extraction success

# Results – Extraction Accuracy by Airline Model & Prompt

Airline Group	Manual KPIs	Report(s) only				Report(s) + Manual KPIs XLS				Report(s) + Prompt including manual KPIs			
		GPT 4.0	o3-mini	R1	Avg	GPT 4.0	o3-mini	R1	Avg	GPT 4.0	o3-mini	R1	Avg
Aegian	100%	42%	46%	25%	38%	0%	63%	71%	44%	96%	71%	79%	82%
Air Baltic	100%	40%	80%	60%	60%	40%	60%	60%	53%	40%	100%	80%	73%
AF-KLM	100%	15%	23%	15%	18%	12%	12%	4%	9%	35%	38%	46%	40%
easyJet	100%	33%	17%	33%	28%	92%	17%	25%	44%	83%	33%	33%	50%
Finnair	100%	35%	35%	25%	32%	0%	100%	85%	62%	90%	90%	90%	90%
IAG	100%	29%	29%	21%	26%	46%	43%	46%	45%	82%	86%	100%	89%
Icelandair	100%	33%	33%	33%	33%	0%	0%	0%	0%	67%	75%	50%	64%
Jet2	100%	17%	33%	33%	28%	17%	17%	0%	11%	100%	33%	33%	56%
Lufhansa	100%	54%	62%	54%	56%	62%	62%	62%	62%	92%	92%	92%	92%
Norse	100%	0%	0%	0%	0%	0%	0%	0%	0%	100%	60%	60%	73%
Norwegian	100%	25%	25%	6%	19%	6%	0%	13%	6%	44%	38%	31%	38%
Ryanair	100%	30%	20%	30%	27%	20%	10%	20%	17%	10%	90%	40%	47%
SAS	100%	33%	56%	17%	35%	50%	50%	44%	48%	100%	56%	94%	83%
TUI	100%	20%	40%	0%	20%	0%	0%	0%	0%	100%	100%	100%	100%
Wizzair	100%	33%	13%	13%	20%	67%	13%	13%	31%	93%	93%	27%	71%
<b>Total</b>	<b>100%</b>	<b>29%</b>	<b>32%</b>	<b>22%</b>	<b>28%</b>	<b>31%</b>	<b>35%</b>	<b>35%</b>	<b>34%</b>	<b>71%</b>	<b>65%</b>	<b>63%</b>	<b>66%</b>

Difference among airlines could be an indication of **more complex** or **less standardized** reporting formats.

# Results – Impact of Data Source Format



**HTML** format **outperformed PDF** for extraction accuracy



**iXBRL** (based on HTML) format **promising**

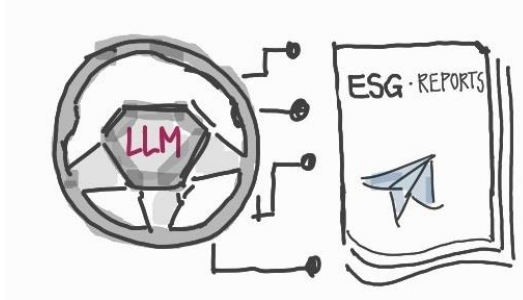


**PDF** remains **common but** is **suboptimal** for automated extraction

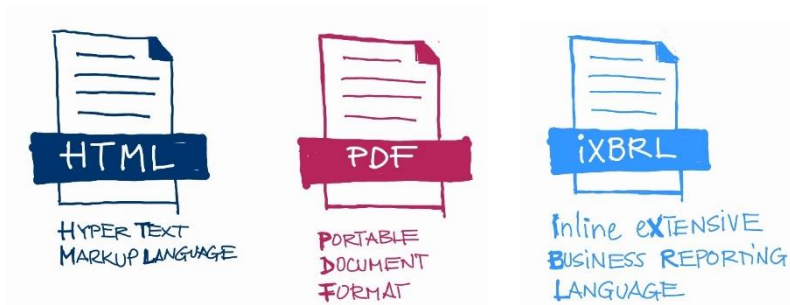
# Conclusions



**LLMs** are a **promising method** for data extraction



**Prompt** is the **dominant factor** in extraction success



**Data format** influences accuracy



Thank you!



[Azero2050.eu](https://Azero2050.eu)



## Enhancing ESG Data Collection: LLM Approaches for Airline Emission KPIs

**Luis MARTÍN-DOMINGO**

Business School, Dublin City University, Dublin 9, Ireland

Faculty of Aviation, Ozyegin University, Istanbul, Turkey.

<https://Azero2050.eu/>



Funded by  
the European Union

This project has received funding from the European Union's HORIZON-MSCA-2023-PF-01 research programme under the Marie Skłodowska-Curie (MSCA) grant agreement No 101151804