do your thing

SecurePrompt

29th September 2025 mAIvericks

What is SecurePrompt ?

Executive Summary

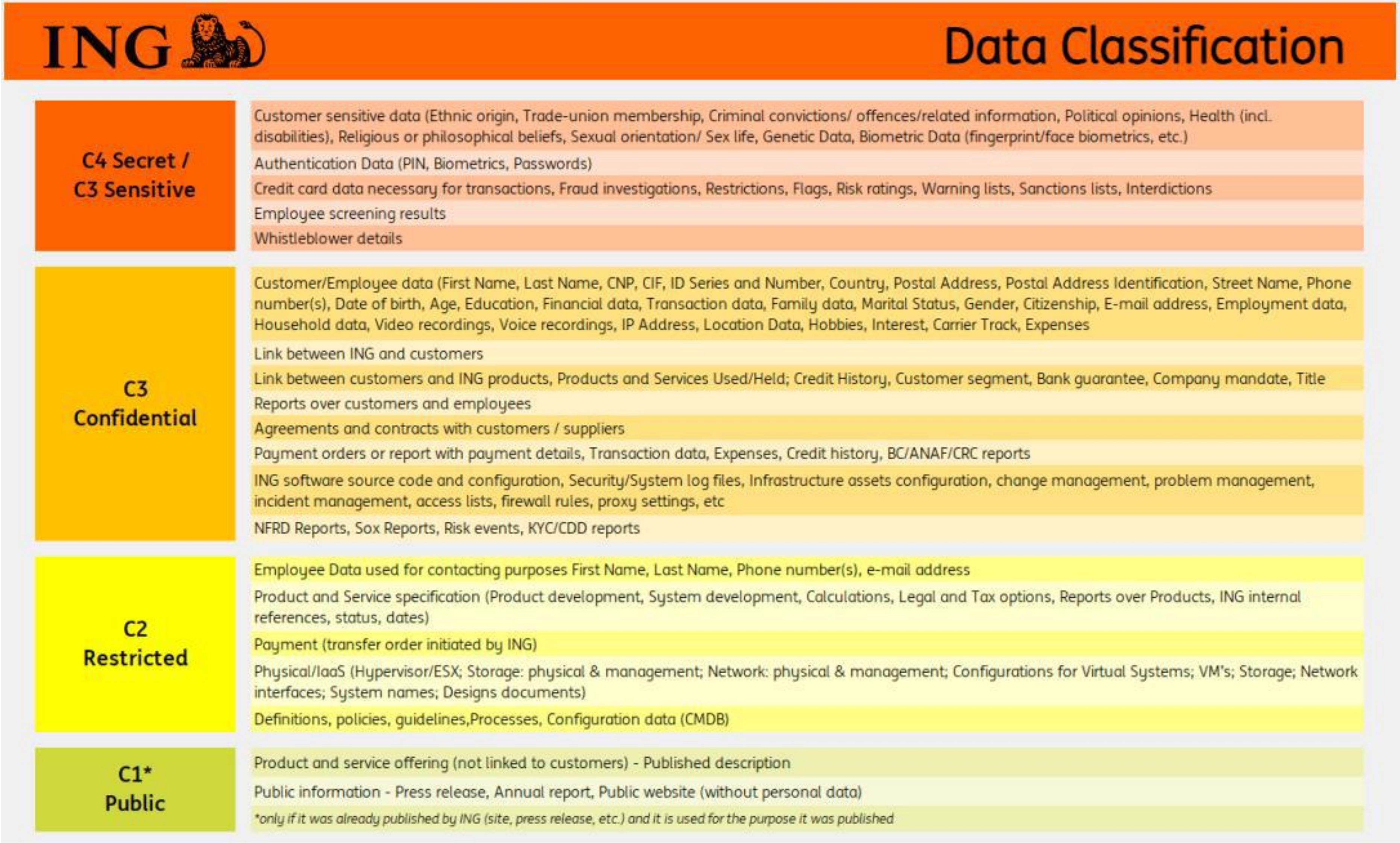
SecurePrompt: a production-grade prompt and file scrubber with controlled de-scrubbing and audit intelligence for banking LLMs

Objective

Develop a secure, auditable, and reversible file / prompt scrubber tailed for use with LLMs in banking. The system must:

- Scrub sensitive data from prompts and files
- Log all transformations into an append-only audit trail
- Allow controlled de-scrubbing of specific entities
- Provide explainability, adaptive sensitivity (C2 – C4), and metrics visibility

ING’s four-tier data classification



ING

Data Classification

C4 Secret / C3 Sensitive	Customer sensitive data (Ethnic origin, Trade-union membership, Criminal convictions/ offences/related information, Political opinions, Health (incl. disabilities), Religious or philosophical beliefs, Sexual orientation/ Sex life, Genetic Data, Biometric Data (fingerprint/face biometrics, etc.)
	Authentication Data (PIN, Biometrics, Passwords)
	Credit card data necessary for transactions, Fraud investigations, Restrictions, Flags, Risk ratings, Warning lists, Sanctions lists, Interdictions
	Employee screening results
	Whistleblower details
C3 Confidential	Customer/Employee data (First Name, Last Name, CNP, CIF, ID Series and Number, Country, Postal Address, Postal Address Identification, Street Name, Phone number(s), Date of birth, Age, Education, Financial data, Transaction data, Family data, Marital Status, Gender, Citizenship, E-mail address, Employment data, Household data, Video recordings, Voice recordings, IP Address, Location Data, Hobbies, Interest, Carrier Track, Expenses
	Link between ING and customers
	Link between customers and ING products, Products and Services Used/Held; Credit History, Customer segment, Bank guarantee, Company mandate, Title
	Reports over customers and employees
	Agreements and contracts with customers / suppliers
	Payment orders or report with payment details, Transaction data, Expenses, Credit history, BC/ANAF/CRC reports
	ING software source code and configuration, Security/System log files, Infrastructure assets configuration, change management, problem management, incident management, access lists, firewall rules, proxy settings, etc
	NFRD Reports, Sox Reports, Risk events, KYC/CDD reports
C2 Restricted	Employee Data used for contacting purposes First Name, Last Name, Phone number(s), e-mail address
	Product and Service specification (Product development, System development, Calculations, Legal and Tax options, Reports over Products, ING internal references, status, dates)
	Payment (transfer order initiated by ING)
	Physical/IaaS (Hypervisor/ESX; Storage: physical & management; Network: physical & management; Configurations for Virtual Systems; VM's; Storage; Network interfaces; System names; Designs documents)
	Definitions, policies, guidelines,Processes, Configuration data (CMDB)
C1* Public	Product and service offering (not linked to customers) - Published description
	Public information - Press release, Annual report, Public website (without personal data)
	**only if it was already published by ING (site, press release, etc.) and it is used for the purpose it was published*

Feature 1:
Prompt Scrubbing

Detect and redact sensitive information.

Feature 2:
File Scrubbing

Return redacted versions of file containing confidential information.

Feature 3:
Audit Logging

Log prompts and scrubbing operations for audit purposes.

Feature 4:
De-Scrubbing

Allow authorized users to reintroduce scrubbed data.

Feature 5:
Scrubbing Sensitivity

Dynamically adjust scrubbing strictness based on parameters.

Bonus:
Scrubbing Dashboard

Track, visualize and predict future scrubbing operations

Feature 1:
Prompt Scrubbing

“Give me all transactions initiated by Ion Postaru, associated to his business account. The credit card linked to that account has the number 0134 5678 9101.”

⇩

“Give me all transactions initiated by customer *associated to his business account. The credit card linked to that account has the number* Credit card *”*

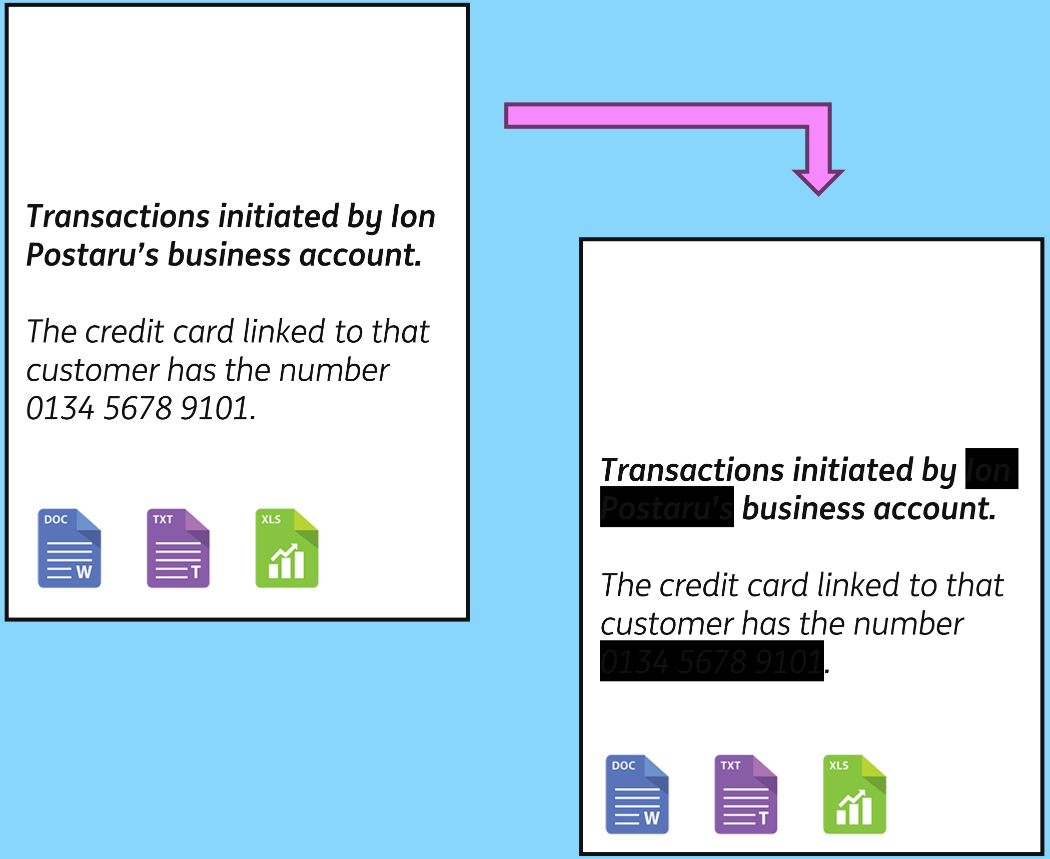
⇩

“Give me all transactions initiated by Customer (78%) *ssociated to his business account. The credit card linked to that account has the number* Credit card (99%) *”*

Detect and redact sensitive information.

- **Detect and redact sensitive information (C2-C4)**
  - **Phase 1**: static anonymization using regex, reference lists, etc.
  - **Phase 2**: anonymization intelligence – a model that predicts data as being confidential, assigns a **class** and **score**
- **Provide an explanation for each redaction**
  - “detected as IBAN”
- **Include confidence scores**
- **Replace scrubbed data with an identifier**

Feature 2:
File Scrubbing



Transactions initiated by Ion Postaru's business account.

The credit card linked to that customer has the number 0134 5678 9101.

Transactions initiated by* [redacted] *business account.

The credit card linked to that customer has the number [redacted].

Return redacted versions of file containing confidential information.

- **Accept files in multiple formats**
  - Pdf, docx, txt, html
  - **Highest priority**: screenshots (.png) formats

- **Follow the same extraction / scrubbing principle as prompt scrubbing**
  - Identify / predict sensitive information
  - Redact and justify

- **Return a redacted version of the file with data being replaced or removed**

Feature 3:
Audit Logging

- **Append-only logs of operations**
  - Timestamp
  - Corporate Key
  - Logon & Logoff actions
  - Opening & Closing session
  - Client identification
    - Device identification
    - Browser identification
    - Location identification (MAC)
  - Original & Scrubbed Content
  - Highlight malicious prompts
    - Search on customer name
    - Injection of data for a customer
    - Update of the customer’s information
    - Deletion of customer information
    - Share / export of customer information
  - Scrubbing actions & metadata
  - Confidence levels

Log prompts and scrubbing operations for audit purposes.

Feature 4: De-Scrubbing

“Give me all transactions initiated by customer *associated to his business account. The credit card linked to that account has the number* Credit card *”*

“Give me all transactions initiated by Ion Postaru, associated to his business account. The credit card linked to that account has the number 0134 5678 9101.”

Allow authorized users to reintroduce scrubbed data.

- **Allow authorized users to request reintroduction of scrubbed data**
- **Provide support for**
  - Full de-scrubbing
  - Selective entity restoration
- **Log all de-scrubbing requests & actions**
- **Nice-to-have**
  - Request justification for de-scrubbing sensitive data

Feature 5: Scrubbing Sensitivity

“Give me all transactions initiated by Ion Postaru, associated to his business account. The credit card linked to that account has the number 0134 5678 9101.”

Parameter: only C4

“Give me all transactions initiated by Ion Postaru, associated to his business account. The credit card linked to that account has the number Credit card *”*

Dynamically adjust scrubbing strictness based on parameters.

- **Dynamically adjust scrubbing strictness based on prompt context (or parameters)**
- **Example parameters**
  - Risk level: C2, C3, C4
  - Labels
    - Customer name
    - IBAN
    - Address
    - …

Bonus:
Scrubbing Dashboard

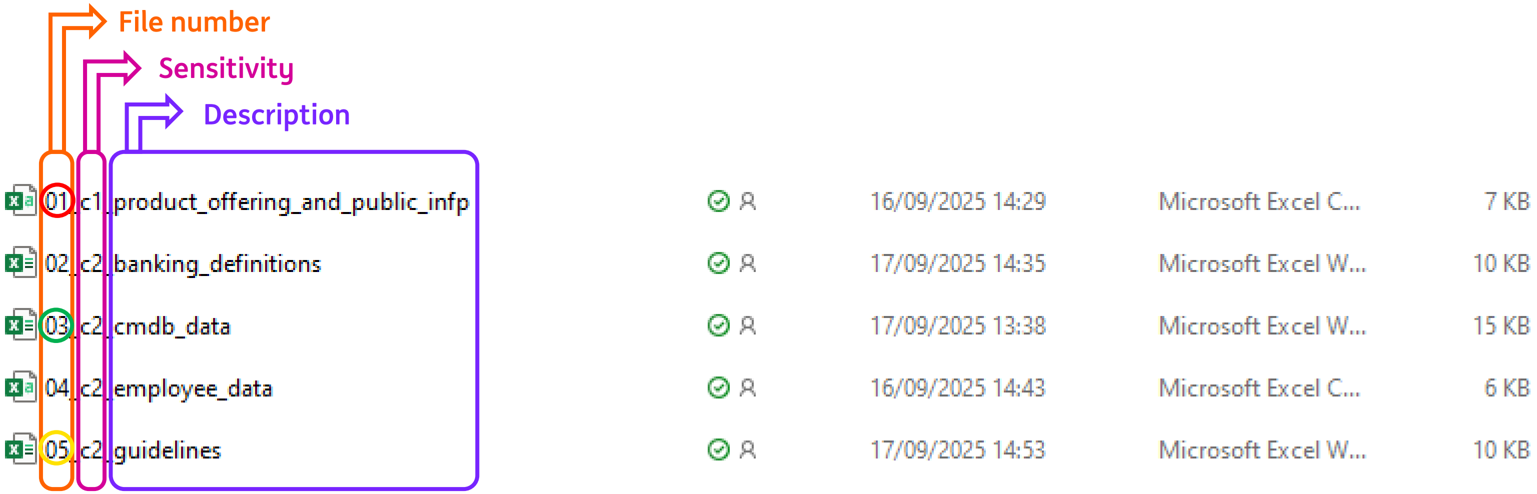
Dynamically adjust scrubbing strictness based on parameters.

- **Tracking and visualization**
  - Volume of scrubbed prompts
  - Types and volume of identified entities
  - Volume of de-scrubbing requests
  - Etc. (impress us …)

- **Insights into systems usage and trends**
  - show usage trends
  - Predict future requests volume

Data Structure

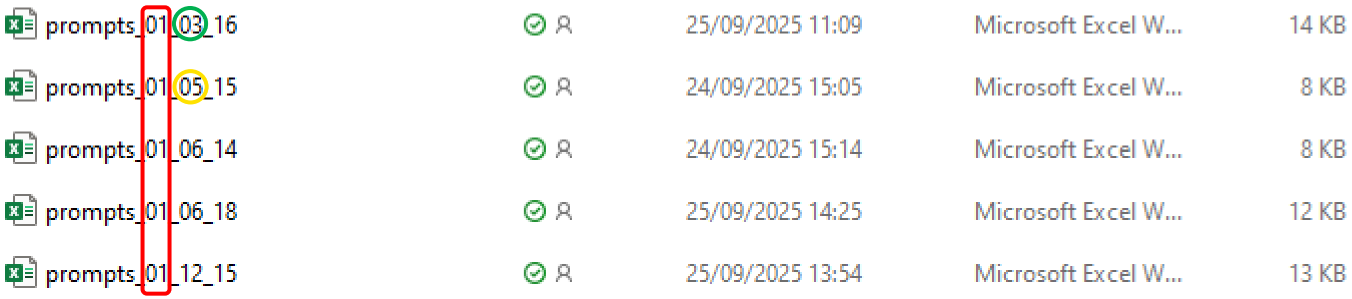
Raw data



File number
Sensitivity
Description

01_c1_product_offering_and_public_infp	✓	16/09/2025 14:29	Microsoft Excel C...	7 KB
02_c2_banking_definitions	✓	17/09/2025 14:35	Microsoft Excel W...	10 KB
03_c2_cmdb_data	✓	17/09/2025 13:38	Microsoft Excel W...	15 KB
04_c2_employee_data	✓	16/09/2025 14:43	Microsoft Excel C...	6 KB
05_c2_guidelines	✓	17/09/2025 14:53	Microsoft Excel W...	10 KB

Prompts

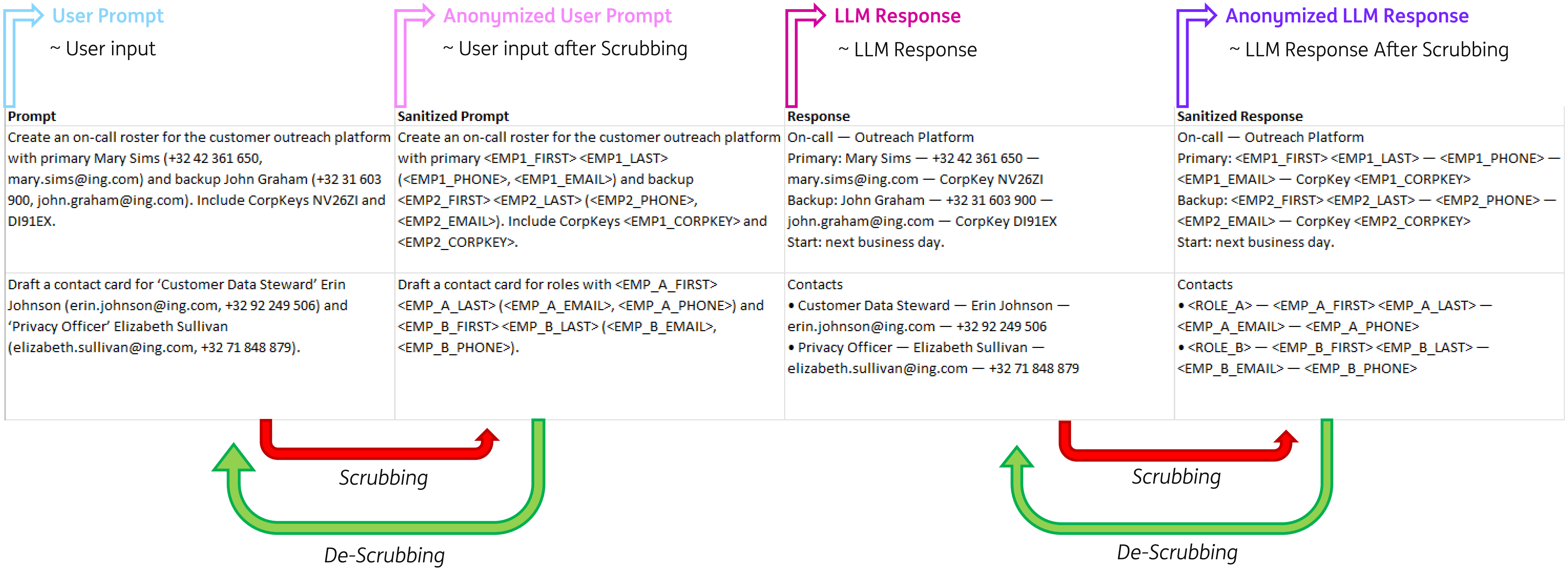


prompts_01_03_16	✓	25/09/2025 11:09	Microsoft Excel W...	14 KB
prompts_01_05_15	✓	24/09/2025 15:05	Microsoft Excel W...	8 KB
prompts_01_06_14	✓	24/09/2025 15:14	Microsoft Excel W...	8 KB
prompts_01_06_18	✓	25/09/2025 14:25	Microsoft Excel W...	12 KB
prompts_01_12_15	✓	25/09/2025 13:54	Microsoft Excel W...	13 KB

Note

These prompts are a "starting point". Feel free to create your prompts, or anonymize differently

Prompt File Contents



Prompt	Sanitized Prompt	Response	Sanitized Response
Create an on-call roster for the customer outreach platform with primary Mary Sims (+32 42 361 650, mary.sims@ing.com) and backup John Graham (+32 31 603 900, john.graham@ing.com). Include CorpKeys NV26ZI and DI91EX.	Create an on-call roster for the customer outreach platform with primary <EMP1_FIRST> <EMP1_LAST> (<EMP1_PHONE>, <EMP1_EMAIL>) and backup <EMP2_FIRST> <EMP2_LAST> (<EMP2_PHONE>, <EMP2_EMAIL>). Include CorpKeys <EMP1_CORPKEY> and <EMP2_CORPKEY>.	On-call — Outreach Platform Primary: Mary Sims — +32 42 361 650 — mary.sims@ing.com — CorpKey NV26ZI Backup: John Graham — +32 31 603 900 — john.graham@ing.com — CorpKey DI91EX Start: next business day.	On-call — Outreach Platform Primary: <EMP1_FIRST> <EMP1_LAST> — <EMP1_PHONE> — <EMP1_EMAIL> — CorpKey <EMP1_CORPKEY> Backup: <EMP2_FIRST> <EMP2_LAST> — <EMP2_PHONE> — <EMP2_EMAIL> — CorpKey <EMP2_CORPKEY> Start: next business day.
Draft a contact card for 'Customer Data Steward' Erin Johnson (erin.johnson@ing.com, +32 92 249 506) and 'Privacy Officer' Elizabeth Sullivan (elizabeth.sullivan@ing.com, +32 71 848 879).	Draft a contact card for roles with <EMP_A_FIRST> <EMP_A_LAST> (<EMP_A_EMAIL>, <EMP_A_PHONE>) and <EMP_B_FIRST> <EMP_B_LAST> (<EMP_B_EMAIL>, <EMP_B_PHONE>).	Contacts • Customer Data Steward — Erin Johnson — erin.johnson@ing.com — +32 92 249 506 • Privacy Officer — Elizabeth Sullivan — elizabeth.sullivan@ing.com — +32 71 848 879	Contacts • <ROLE_A> — <EMP_A_FIRST> <EMP_A_LAST> — <EMP_A_EMAIL> — <EMP_A_PHONE> • <ROLE_B> — <EMP_B_FIRST> <EMP_B_LAST> — <EMP_B_EMAIL> — <EMP_B_PHONE>

Milestones, Project & Performance

03/10 - Milestone 1: intermediary presentation + Q&A, online

- Performance measure: % of correctly identified and scrubbed entities
- Prepare your questions accordingly on forehand
- A short slide deck is always welcome

10/10 - Milestone 2: final presentation @ ING office

- Performance measure: pass / fail
  - all sensitive information must be scrubbed according to the set parameters to be considered as successful.
- A slide deck showcasing performance, strengths, weaknesses and potential improvements is expected

Good to know

For questions and clarifiactions, please send a mail to **tanguy.dhose@ing.com** – be clear, specific and reasonable.

Deliverables

Specifications

- The project should be implemented in Python 3
- Use the modules you want, if they are free and open source
- The project is delivered as one or multiple modules (object oriented)
  - Be able to explain your choices
- Each significant method is covered by unit tests to:
  - Guarantee the success rate of scrubbing
  - Showcase the precision of your modules
- Delivered modules should strive for maximal efficiency
  - Speed of execution is a critical measure of success
- Extensive documentation is expected
  - Regular single line comments to clarify operations
  - A detailed docstring for each function or method, providing the necessary information on the input, operations and return value