

CVoria Privacy Model Whitepaper

Removing personal details from CVs before AI analysis



Published 2026-10-08

23 / 757
CVoria model

108 / 757
GLiNER multi PII

199 / 757
LiquidAI PII

299 / 757
Microsoft Presidio

Core personal details left readable in 200 new fictional CVs. Lower is better.

Hosted LLMs given CVoria's policy: Gemini 3.8 Flash 7, GPT-5.6 Luna 17 (Section 5).

Target system: the CVoria privacy model, a 307M-parameter mmBERT-base token classifier, run with the code and model file deployed on AWS Lambda in Stockholm.

Open tools compared: Microsoft Presidio, GLiNER multi PII and LiquidAI LFM2.5 PII (350M).

Hosted LLMs compared: Gemini 3.8 Flash and GPT-5.6 Luna, each given CVoria's redaction policy.

Test data: 200 fictional CVs (150 excerpts, 50 full CVs) written by hand for this benchmark by two AI models, labelled twice independently, and locked before any system was run.

Benchmark type: how many personal details stay readable after each system's output is blanked out of the text.

Headline result: On 200 new fictional CVs, the CVoria privacy model left 23 of 757 core personal details readable (3.0%): names, email addresses, phone numbers, street addresses, personal links, dates of birth and ID numbers. GLiNER multi PII left 108, LiquidAI LFM2.5 PII 199 and Microsoft Presidio 299. Given CVoria's written policy, Gemini 3.8 Flash left 7 and GPT-5.6 Luna 17, but a hosted model has to receive the full CV in order to remove the details.

Contents

1	Executive Summary	3
2	Why CVoria Removes Personal Details First	3
3	The CVoria Privacy Model	4
4	Benchmark Design	4
4.1	A new test set	4
4.2	Systems	5
4.3	How readable details are counted	6
5	Results	7
5.1	All systems	7
5.2	Under each labeller's answer key	7
5.3	By type	8
5.4	By kind of CV, language and writer	8
5.5	Precision and over-redaction	9
6	Limitations and Interpretation Boundary	9
7	Reproducing the Results	10
8	Next Steps	10

1 Executive Summary

Before CVoria sends CV text to an AI model, its own privacy model replaces the personal details it detects with placeholders. This whitepaper measures how well that model does it, compared with three widely used open tools and two hosted large language models.

The main measure is simple: after each system's output is blanked out of the CV, how many personal details from the answer key can still be read? The test set is new: 200 fictional CVs written for this benchmark, checked against the model's training data, labelled twice independently and locked, with checksums, before any system was run.

The CVoria privacy model left **23 of 757** core personal details readable. GLiNER multi PII left 108, LiquidAI LFM2.5 PII 199 and Microsoft Presidio 299. The CVoria model left no phone number and no ID number readable; 15 of its 23 misses were names, mostly in forms such as children's first names, nicknames and the candidate's own name written as a citation.

Hosted LLMs given CVoria's written policy did better: Gemini 3.8 Flash left 7 core details readable and GPT-5.6 Luna 17, and both were more precise. Using them for this step would mean sending the full CV, personal details included, to the provider. The CVoria model runs without a GPU in CVoria's own AWS account in Stockholm, so the details it detects are replaced before the text reaches an AI provider.

Main claim supported by this benchmark: On 200 new fictional CVs, CVoria's privacy model left far fewer personal details readable than Microsoft Presidio, GLiNER multi PII and LiquidAI LFM2.5 PII. Gemini 3.8 Flash and GPT-5.6 Luna, given the same policy, left fewer still, but only after receiving the full CV.

The benchmark has clear limits. The CVs are fictional and were written by AI models, and they were labelled under the same policy the CVoria model was trained on. Real CVs are longer and messier, and results on them may be worse.

2 Why CVoria Removes Personal Details First

CVoria uses large language models from external providers for CV analysis, cover letters, job matching and interview practice. A CV is full of personal data, and most of the analysis does not need it: the feedback on a work history does not depend on the candidate's name, phone number or home address.

CVoria therefore runs CV text through its own privacy model before the external AI calls that use it:

- The model looks for the personal details in the CV text.
- Each detail it finds is replaced, everywhere it occurs, with a numbered placeholder such as [NAME_1] or [EMAIL_1]. The same value always gets the same placeholder.
- The redacted copy is sent to the AI provider. Details the model misses are sent as written.
- Where the result needs the real details back, for example the signature of a cover letter, CVoria puts them back into the AI's answer itself.
- If the privacy model cannot be reached, the request is refused instead of being sent unredacted.

The model runs on AWS Lambda in the Stockholm region, on CPU only, in CVoria's own AWS account.

3 The CVoria Privacy Model

The model is a token classifier built on mmBERT-base, a multilingual encoder with 307 million parameters. It reads the CV text and tags every word as either part of a personal detail of one of 13 types, or not. Whole words are tagged, and a detail found once is replaced everywhere it occurs in the CV.

The 13 types follow CVoria’s redaction policy, which is included in the public data package:

Type	What it covers
PERSON	The candidate’s name and other private individuals
EMAIL, PHONE	Personal contact details
ADDRESS	Street and postal addresses (not a city on its own)
PERSONAL_URL	Personal links such as LinkedIn or GitHub profiles
DATE_OF_BIRTH, GOV_ID	Birth dates and ages, identity, tax and passport numbers
REFERENCE	A referee’s name and contact details
IMMIGRATION_STATUS	Visa, residence and work permit status
PROTECTED_ATTRIBUTE	Health, religion, ethnic origin, union membership and similar
CERTIFICATE_ID, BANK_OR_PAYMENT	Certificate and licence numbers, bank and payment details
SIGNATURE	A name used as a signature

The model was trained on about 6,600 generated CVs plus 1,400 reformatted copies of them (for example with letter-spaced names or broken lines), mostly in English with a substantial share in Swedish, labelled under the policy. No real person’s CV was used for training. In production the model runs with an int8 embedding table, which halves the model file from 1.23 GB to 641 MB; on this benchmark the full-precision model left 24 core details readable against 23 for the served one. Long CVs are read in overlapping 3,000-character chunks.

4 Benchmark Design

4.1 A new test set

CVoria’s earlier internal tests used hand-written CV excerpts that shared some personal details with the training data and had been used to choose the model’s decoding method and confidence threshold. Other design choices, such as the chunk size and the training data mix, were made on separate internal development data. Results on such data overstate how the model does on CVs it has never seen. For this whitepaper a new set was written and fixed before any system ran on it:

- **Written by hand by two models.** Claude Opus 5.5 and GPT-6.1 Sol each wrote 100 CVs, at medium reasoning effort, with no tools: every CV was typed directly into the model’s reply, not produced by a program. Each worked from the same brief and its own list of 100 CV descriptions (jobs, countries, languages and the personal details to include), all published in the package. Neither saw the training data or any earlier test set.
- **Fictional contact details.** Phone numbers come only from ranges reserved for fiction in Sweden, the UK and the US. Every email and website domain was looked up and did not exist when checked, GitHub handles were checked to be free or invalid, US social security

numbers are in the range reserved for advertising, and Swedish personal identity numbers fail their check digit. Profile names on other platforms, such as LinkedIn, are invented but could not be checked.

- **Independent of training.** Every answer-key value of six characters or more and every line of 40 characters or more was searched for in all training files, earlier test sets and the policy. No line matched. The matching values were phone numbers, which come from the small ranges reserved for fiction, ordinary words such as nationalities, one first name and one birth date; one certificate number that matched was changed before the set was locked.
- **Labelled twice.** Each model labelled the other’s CVs without seeing the writer’s answer key. The two keys agreed on 980 of 1,091 items (89.8%). CVoria settled the 111 differences by the policy’s rules, mostly about where a value starts and ends. The package lists every difference and every decision, and Section 5.2 shows the results under each labeller’s key.
- **Locked first.** The test set, the prompt input, the policy and the scoring script were fixed and their SHA-256 checksums recorded before any system was run. Each system then gave one answer per CV. The checksums are local records: nothing outside CVoria attests when they were written.

Two changes were made after the lock, both recorded with new checksums in the package’s `LOCK.md`. A Stack Overflow profile link in one CV contained a numeric user ID that could have opened a real person’s profile; the ID was replaced with 0000000 in the CV, the answer keys and every system’s output, which changed no result. And the hosted LLMs are scored on their own answers rather than on the run script’s post-processed version, which had dropped some spellings of a name as duplicates without regard to case and so left them readable; this lowered their counts from 12 to 7 for Gemini 3.8 Flash and from 21 to 17 for GPT-5.6 Luna.

Part	CVs	Core details	Description
Excerpts	150	460	Parts of CVs (median 548 characters): headers, references, personal details, publication lists, messy PDF extraction
Full CVs	50	297	Complete CVs (median 4,055 characters), about one in six written as text extracted from two-column PDFs
Swedish	72	212	German, French, Spanish, Norwegian, Polish, Finnish and others
English	108	444	
15 other languages	20	101	

The set has 1,041 personal details in all, 757 of them core details.

4.2 Systems

- **CVoria privacy model:** the model file and redaction code deployed on AWS Lambda, run locally with the same settings.
- **Microsoft Presidio:** presidio-analyzer 2.2.364 with every default recognizer and spaCy `en_core_web_lg` 3.8.0, an English model, on all CVs, including the 92 that are not in English. Presidio has no recognizer for street addresses. Without its date recognizer, the configuration used in CVoria’s earlier tests, it left 325 core details readable instead of 299.
- **GLiNER multi PII:** `urchade/gliner_multi_pii-v1`, zero-shot with 20 plain-language labels such as “person”, “phone number” and “national id number” (all listed in the package), threshold 0.5, on 1,500-character chunks.

- **LiquidAI LFM2.5 PII (350M):** LiquidAI/LFM2.5-Encoder-350M-PII-Detector with its published hybrid decoder. Company names and money amounts were dropped from its output.
- **Gemini 3.8 Flash and GPT-5.6 Luna:** hosted LLMs given a system prompt, CVoria’s full written redaction policy and the CV, with medium reasoning effort, run through command-line tools on CVoria’s computer. Gemini ran through the Antigravity CLI, which has no structured output, so the prompt asked for a JSON reply and no tool use; ten of its 200 requests failed in the tool and were sent again once. GPT ran through the Codex CLI with a JSON output schema. Both tools add their own system instructions.

The open tools were not tuned for CVs.

4.3 How readable details are counted

For every system, items with fewer than three letters or digits are dropped from its output, as CVoria does in production. Every occurrence of every remaining item is then blanked out of the CV text; any run of whitespace matches any other, and a match may not start or end inside a word. A personal detail from the answer key counts as **readable** if, in any place it occurs, at least three of its letters or digits are still visible (or all of them, for shorter values). The rule was fixed before any system was run and has no list of exempt words.

Core personal details are names, email addresses, phone numbers, street addresses, personal links, dates of birth and government ID numbers: the details that identify or reach a person. Readability ignores the type a system assigns, so it does not depend on how each tool’s labels were mapped to CVoria’s.

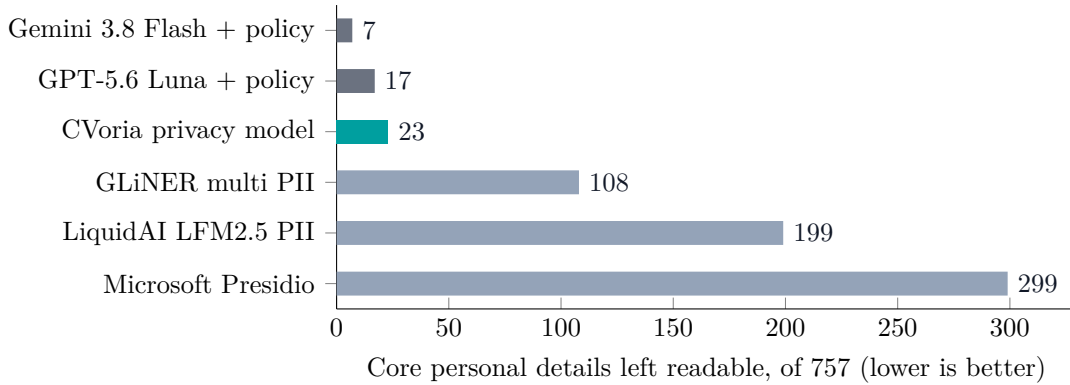
Over-redaction is measured three ways. **Precision** is the share of returned items that overlap a personal detail in the answer key; the type and where a value starts and ends do not matter, so “Tel. 070-174 06 23” for the value “070-174 06 23” counts as correct, while an employment date returned as a birth date does not. Because the overlap is checked wherever the same text occurs, a word that is a personal detail in one place can make an item count as correct elsewhere. **Character precision** is the share of blanked characters that belong to a personal detail, which counts a value that swallows surrounding text against the system. **Other text blanked** is the share of the CV’s other text that was removed. The lock fixed an exact-match precision instead, which needs the same type and text, ignoring case, spacing and punctuation. It counted items like “Tel. 070-174 06 23” as wrong, so precision and character precision were added after the first results were seen; the exact-match figures are in Section 5.5 and the package. Both added measures are computed the same way for every system and change no readable count.

5 Results

5.1 All systems

System	Core readable	All readable	Precision	Character precision	Other text blanked
CVoria privacy model	23 / 757	70 / 1,041	94.2%	94.4%	0.71%
GLiNER multi PII	108 / 757	318 / 1,041	77.4%	79.8%	2.20%
LiquidAI LFM2.5 PII	199 / 757	438 / 1,041	88.9%	91.2%	0.70%
Microsoft Presidio	299 / 757	516 / 1,041	37.1%	36.9%	13.12%
Gemini 3.8 Flash + policy	7 / 757	24 / 1,041	100.0%	98.2%	0.23%
GPT-5.6 Luna + policy	17 / 757	34 / 1,041	99.1%	97.2%	0.35%

Table 1: Personal details left readable, precision and over-redaction on all 200 CVs. “All readable” counts every type, including references, certificates and protected attributes. The best value in each column is highlighted; the open tools and the CVoria model run locally, the LLMs as hosted services.



5.2 Under each labeller’s answer key

The final answer key settles the disagreements between the two labellers. Scored against the writer’s own key or the blind second key instead, the counts change by a few details and the order of the systems stays the same:

Core details left readable	Writer’s key	Second key	Final key
Gemini 3.8 Flash + policy	11	9	7
GPT-5.6 Luna + policy	15	19	17
CVoria privacy model	26	24	23
GLiNER multi PII	109	107	108
LiquidAI LFM2.5 PII	199	199	199
Microsoft Presidio	297	297	299

Table 2: Core details left readable under each answer key.

5.3 By type

Type (of total)	CVoria	GLiNER	LiquidAI	Presidio	Gemini	GPT
Names (225)	15	13	87	93	5	8
Email addresses (119)	1	6	5	5	0	0
Phone numbers (132)	0	10	12	53	0	0
Street addresses (100)	2	20	30	100	0	6
Personal links (82)	3	15	50	5	0	1
Dates of birth (36)	2	2	3	10	2	2
ID numbers (63)	0	42	12	33	0	0
Health, religion etc. (72)	27	37	66	43	14	11
Immigration status (21)	10	20	21	21	1	1
References (90)	9	90	90	89	1	2
Certificate numbers (73)	0	55	54	46	0	3
Bank and payment (11)	0	6	3	10	1	0
Signatures (17)	1	2	5	8	0	0

Table 3: Readable details by type: the seven core types, then the others.

The CVoria model left no phone number or ID number readable. Of its 15 name misses, four were the candidate’s own name written as a citation (“Kothapalli, V.”), three were children’s first names, three were nicknames or call signs, two were surnames left visible next to a removed first name, two were the candidate’s first name or surname written on its own, and one was a name in vCard format; GLiNER left slightly fewer names readable (13). The model’s other core misses were two ages (“Age: 34”), three personal links (a GitLab and a Dribbble profile and a personal website), part of a letter-spaced email address, and a city and a country left next to a removed street address.

Outside the core types, the CVoria model is weakest on health, religion and similar details (27 of 72 readable) and on immigration status (10 of 21), where both hosted LLMs did much better. A reference counts as readable if three letters or digits of it remain, including the referee’s role and employer; the open tools do not treat those as personal data, which explains most of their 89 to 90 of 90.

5.4 By kind of CV, language and writer

Subset (core details)	CVoria	GLiNER	LiquidAI	Presidio	Gemini	GPT
Excerpts (460)	16	71	110	179	6	8
Full CVs (297)	7	37	89	120	1	9
Swedish (212)	5	19	72	116	1	7
English (444)	18	69	119	135	6	10
Other languages (101)	0	20	8	48	0	0
Written by Claude (428)	19	72	130	170	5	8
Written by GPT (329)	4	36	69	129	2	9

Table 4: Core details left readable, by subset.

In every subset the CVoria model leaves several times fewer core details readable than the best open tool, and Gemini 3.8 Flash leaves the fewest. GPT-5.6 Luna leaves fewer than the CVoria model in excerpts and English CVs, and more in full CVs and Swedish CVs. As shares of each subset’s core details, the CVs Claude Opus 5.5 wrote were harder for every system except

Presidio, which was level (39.7% and 39.2%), and GPT-5.6 Luna, which left more readable in the CVs written by GPT-6.1 Sol (2.7%) than in Claude’s (1.9%). There is no sign that it benefits from writing by a model from the same company.

5.5 Precision and over-redaction

Of the 1,122 items the CVoria model returned, 65 (5.8%) overlap no personal detail in the answer key: 17 dates (employment, issue, expiry and signing dates, 12 of them returned as birth dates), 11 details of an employer such as its address, website, phone number or company number, 10 headings or labels such as “LEBENSLAUF” or “Referens”, 10 legal, course, patent or document numbers, 6 place names, 4 public figures, 2 others, and 5 that the answer key arguably should have included: relationship and marital status, union and professional-association membership, and a file name containing the candidate’s name. The dates matter most, because the AI analysing the CV then loses part of the candidate’s timeline. Gemini 3.8 Flash returned nothing outside the answer key and GPT-5.6 Luna 9 items.

The exact-match precision fixed in the lock gives 75.8% for the CVoria model, 96.0% for Gemini 3.8 Flash, 93.3% for GPT-5.6 Luna, 53.3% for GLiNER, 46.6% for LiquidAI and 11.7% for Presidio. The CVoria model loses most there because it often returns a value in parts, for example a reference as separate name, phone and email items instead of one reference, or with different boundaries than the answer key. Presidio blanked 13.1% of the text that is not personal data, largely because it also removes dates.

6 Limitations and Interpretation Boundary

- **Fictional data written by AI models.** No real CV is part of this benchmark. The CVs were written by Claude Opus 5.5 and GPT-6.1 Sol and may differ from real CVs in style and in how densely they pack personal details. Real CVs are longer and messier, and results on them may be worse, in particular more over-redaction.
- **Home advantage.** The answer keys follow the same policy the CVoria model was trained on, and CVoria settled the disagreements between the two labellers. The open tools follow their own, broader or narrower, ideas of personal data and were not tuned for CVs.
- **Same model family.** GPT-6.1 Sol wrote half the CVs and labelled the other half, and GPT-5.6 Luna is from the same provider. Results by writer are reported above.
- **Answer keys.** Two independent labellers and a review of every disagreement leave errors the two labellers share. One found after the set was locked: a file name containing a candidate’s name (CV s042) is not in the key. It was left unchanged to keep the lock.
- **One value too short to remove.** The scorer drops returned items with fewer than three letters or digits, so the only shorter answer-key value, the age “34” in CV o097, counts as readable for every system, including Gemini 3.8 Flash, which returned it.
- **Other people in a CV.** CVoria’s policy does not remove the names of co-authors in publication lists or of public figures. A referee’s name and contact details are removed; the answer keys give the whole reference, including the referee’s role and employer, its own type, counted under all details rather than core details.
- **What is measured.** The benchmark measures how many personal details stay readable before text is sent to an AI model. It does not measure the quality of the analysis that

follows, and the redacted text is still personal data: work history, employers and dates stay in it.

Safest public claim: On 200 new fictional CVs, CVoria’s own privacy model left far fewer personal details readable than Microsoft Presidio, GLiNER multi PII and LiquidAI LFM2.5 PII. Hosted LLMs given the same policy left fewer still, but only after receiving the full CV; CVoria’s model replaces the details it detects in CVoria’s own AWS account in Stockholm, before the text reaches an AI provider.

7 Reproducing the Results

The public data package contains every test CV with its answer key, the prompt input the hosted LLMs were given, every system’s output, both labellers’ keys with every disagreement and decision, the writing brief and CV lists, the redaction policy and the scoring scripts. The scripts need Python 3.9 or newer and nothing else. From the package folder:

```
python scripts/verify_results.py
```

The script checks that the test set, prompt input, policy and scoring script match the checksums in `LOCK.md`, recalculates every table in this whitepaper from the saved outputs, including the results under each labeller’s key, and ends with:

```
All results match results/results.json.
```

To score one system and list every detail it left readable (on Windows, first set the environment variable `PYTHONIOENCODING` to `utf-8`):

```
python scripts/score.py data/test_set/handwritten_v4.jsonl data/predictions/cvoria.jsonl --list
```

The package rescores the saved outputs; it does not rerun the systems, and it does not recalculate the descriptive figures in the text, such as medians, language counts and the descriptions of individual misses, which can be checked against the test set by hand. The CVoria model is not public. The open tools and hosted models can be rerun with the configurations described above and listed in `results/results.json`.

8 Next Steps

The next model update will stop removing employment, issue and expiry dates, which the model sometimes mistakes for birth dates, and handle the name forms and the health and immigration details it missed here. This set has now been used to find those errors, so the update will be tested on a new fictional set, written and locked the same way before any system runs.