

Model Assurance Attestation

Issued 2026-10-15T03:08:56Z · Template 0.1

katuni4ka/tiny-random-chatglm2

aae760a874f44240faefc3f21b3907b7379822ef6fb1c38baa64b336c25538f2
model directory · document SOCAIR-20261015-aae760a8 · issuer socair.ai examples
no license stated · declared base katuni4ka/tiny-random-chatglm2

Tier 1 (static)

Promotion: withheld

3 pass 0 fail 1 lead 1 not tested

CHECKS PERFORMED

Check	Looks for	Result	Evidence
-------	-----------	--------	----------

Format and structure

Malformed container structure, unexpected tensors

PASS means: The container parsed, and its tensor data tiles the data section exactly: no unaccounted bytes, overlapping tensors, or duplicate names. For an ONNX model: every tensor holds exactly the data its type and shape need, no two initializers of a graph share a name, external data stays inside the model's directory and fits its file, and every operator is ONNX's own, one of onnxruntime's contrib operators, or a function the model defines. For an OpenVINO IR: every Const lies inside the .bin and is exactly the size its element type and shape take, the Consts cover the .bin with no byte outside them, and every operation OpenVINO builds is in its own opsets. Tensor values are not judged, and an operator library a deployment registers is outside the model. For a Keras model (HDF5 or a .keras archive): every dataset's storage lies inside the HDF5 data, sized by its type and shape, with none overlapping and no data or link in another file, the archive holds only what Keras reads, and the config names no Python bytecode, pickled module, or class or function from outside Keras and TensorFlow (KerasHub, KerasCV, and KerasNLP classes are named). Weight values are not judged. Addresses AML.T0010.003, LLM03

PASS

1 of 1 file(s) PASS (model.safetensors);
e.g. model.safetensors: safetensors layout
verified: 18 tensors tile 77303328 data
bytes exactly, each sized to its shape and
dtype; 1 metadata keys

File inventory and payloads	Hidden files, embedded payloads, unexpected executables PASS means: No script, shell command, long encoded blob, or executable signature was found in string metadata values, including every string-array element of 64 bytes or more, and no native executable among the files. Shorter array elements, such as vocabulary tokens, are too short to carry one and are not scanned. An archive among the files counts only when the Pickle opcode scan PASSEd it and read every entry in it: a torch.save checkpoint with nothing outside torch.save's layout. Addresses AML.T0011.000, AML.T0018.002, AML.T0010.003, LLM03	PASS	2 of 2 file(s) PASS (directory listing, model.safetensors); e.g. directory listing: repo: 12 files, 4 scripts inventoried (scripts are expected, not a finding); no pickle file: none is named as one or begins as one (a protocol 2 to 5 stream, or a zip holding a pickle), so no Pickle opcode scan row
Hash, provenance, lineage	Traceable origin: a manifest bound to this hash, from an immutable upstream commit; files that disagree with the uploader's checksum list PASS means: A trusted publisher signature verified over this artifact, or an operator-supplied manifest binds this hash to a repository at an immutable commit. The notes say which; a manifest is the operator's claim, not a signature. No file the scan compared with a SHA-256 list its uploader ships (SHA256SUMS or a .sha256 file) disagrees with it, save Markdown, which is named and not graded; agreement is noted, and the PASS does not rest on it. Addresses AML.T0010.003, AML.T0115.001, LLM03	PASS	origin bound to this artifact's hash, at an immutable commit: repo https://huggingface.co/katuni4ka/tiny-random-chatglm2; revision 44385a0daf9b45ce1ae29917ce480ac31c4f10fc; commit 44385a0daf9b45ce1ae29917ce480ac31c4f10fc; a selection of the repo's files at that commit (--exclude runs/ --exclude training_args.bin), 2 left out: runs/Mar01_07-05-09_6aeb10c326b5/events.out.tfevents.1709276714.6aeb10c326b5.199.0, training_args.bin; via airlock pull; publisher signature not established; manifest provenance.json (sha256 5296b6700c9a1016c44c1862763d3db65315b6968febaf964c329eb8e07c490a)

Tokenizer config	Tampered tokenizer tables: special-token ids, token types, control tokens PASS means: The tokenizer tables are internally consistent, and no control token carries instructions. When a canonical reference table for its family is configured (from a feed or SOCAIR_TOKENIZER_REFER ENCE), every token was compared with it and no ordinary token differs; the notes say which reference, or that none was configured or matched. Addresses AML.T0018.003, AML.T0011.000, AML.T0010.003, LLM03	NOT_TESTED	no tokenizer.json, so the tokenizer the loader builds was not inspected; present: special_tokens_map.json, tokenizer.model, tokenizer_config.json
------------------	--	------------	--

Remote code	Python a loader would import: modules a config names (auto_map, custom_pipelines, sentence-transformers and diffusers modules), files loaders import by their path, and the repository files those import PASS means: No config a loader reads names Python in this directory or in another repository (through an auto_map entry, custom_pipelines, a sentence-transformers module, or a diffusers custom pipeline or component), and no file sits where a loader imports one by its path (custom_generate/generate.py, or an inference toolkit's handler.py, pipeline.py, or code/inference.py). Other Python and script files are listed in the notes and not judged: a loader Socair does not model, or a person, can still run them. Addresses AML.T0011.000, AML.T0010.001, LLM03	LEAD medium	<pre>auto_map: config.json: AutoConfig -> katuni4ka/tiny-random-chatglm2--configurat ion_chatglm.ChatGLMConfig (names code in another repository, katuni4ka/tiny-random-chatglm2, not in this artifact; imported with trust_remote_code); auto_map: config.json: AutoModel -> katuni4ka/tiny-random-chatglm2--modeling _chatglm.ChatGLMForConditionalGenerati on (names code in another repository, katuni4ka/tiny-random-chatglm2, not in this artifact; imported with trust_remote_code); auto_map: config.json: AutoModelForCausalLM -> katuni4ka/tiny-random-chatglm2--modeling _chatglm.ChatGLMForConditionalGenerati on (names code in another repository, katuni4ka/tiny-random-chatglm2, not in this artifact; imported with trust_remote_code); auto_map: config.json: AutoModelForSeq2SeqLM -> katuni4ka/tiny-random-chatglm2--modeling _chatglm.ChatGLMForConditionalGenerati on (names code in another repository, katuni4ka/tiny-random-chatglm2, not in this artifact; imported with trust_remote_code); auto_map: config.json: AutoModelForSequenceClassification -> katuni4ka/tiny-random-chatglm2--modeling _chatglm.ChatGLMForSequenceClassificati on (names code in another repository, katuni4ka/tiny-random-chatglm2, not in this artifact; imported with trust_remote_code); auto_map: tokenizer_config.json: AutoTokenizer[0] -> katuni4ka/tiny-random-chatglm2--tokenizati on_chatglm.ChatGLMTOKENIZER (names code in another repository, katuni4ka/tiny-random-chatglm2, not in this artifact; imported with trust_remote_code)</pre>
-------------	---	-------------	--

Framework editions: MITRE ATLAS 2026.09, OWASP LLM Top 10 2025.

ASSURANCE LEVEL: TIER 1 (STATIC)

What it is

Static checks on the artifact's bytes and metadata, run offline. No inference is run, and the weights are not evaluated for behavior.

What it means

Each listed check ran on the exact bytes identified by the artifact hash, and each row states what its PASS means. A FAIL carries positive evidence, a LEAD is a suspicious signal that needs a person's review, and a NOT_TESTED names why it was not tested. A FAIL or a LEAD withholds promotion, and no acceptance clears it.

What it does not mean

It is not an assessment of the model's behavior or safety. It does not test the weights for backdoors or poisoning, behavior that

appears only after quantization or on particular hardware, or jailbreak susceptibility and harmful capability, and it is not a legal review of the model's license.

Tier 2

Tier 2 (inference on the production node class) did not run.

BOUNDED STATEMENT

For the artifact identified by its hash, the Tier 1 checks found the indicators this report lists: each FAIL is positive evidence, and each LEAD is a suspicious signal that needs review. Rows marked NOT_TESTED were not examined, for the reason each row gives. Tier 1 runs no inference, so no model behavior was tested. Every surface outside that scope is listed under Out of scope.

OUT OF SCOPE

This attestation does not certify the absence of unknown backdoors.

- Backdoors, trojans, or poisoning in the model weights. Tier 1 reads the weights' layout, never their behavior.
- Triggered, sleeper, or polymorphic behavior. Tier 1 runs no inference.
- Behavior that appears only after quantization, or only on particular hardware or serving stacks, unless Tier 2 ran on the production node class.
- Behavioral safety: jailbreak susceptibility, harmful capability, and bias.
- Malicious behavior that only emerges at runtime under real traffic.
- Artifact formats we do not parse.
- Pickle code execution through a reviewed class's state, loader bugs, parse differentials, a tampered environment, or code carried as data for a later stage to run.
- Chat-template instructions written as ordinary guidance (no override or concealment phrase, URL, hidden or obfuscated text, or condition on message content other than a standard content-part type such as text or image): the text a rendered template adds, and the text a template adds for a content-part type, is listed in the report, not judged, unless the template matches a reviewed template.
- Legal review of a license, and compliance with its usage policies. The license is identified, and checked only against a policy you configure.

Checks not run at this level:

- Forward-pass trigger probes (Tier 2) (Behavior under the production serving stack): not run. Tier 2 (forward-pass, on the production node class) did not run; Tier 1 is static and runs no inference
- Serving-stack differential (Tier 2) (Same artifact behaving differently across stacks): not run. Tier 2 (forward-pass, on the production node class) did not run; Tier 1 is static and runs no inference
- Known-bad hash match (Match against the known-bad artifact denylist): not run. no known-bad list was supplied (SOCAIR_DENYLIST, or a signed feed that carries one), so the artifact's hash was not matched against one
- Chat template (Code reach, hidden or obfuscated text, override or content-triggered instructions, and the text the rendered template adds to the prompt): not run. the artifact carries no chat template (none in tokenizer_config.json, chat_template.json, or a .jinja file), so there was none to read; a template a serving stack supplies is outside the artifact

Node classes not tested:

- every node class: Tier 1 is static and ran no inference on any hardware

PROMOTION AUTHORIZATION

Clean-store promotion: withheld · Tier 1 only

Withheld: a suspicious lead on Remote code. A FAIL or a LEAD withholds promotion, and no acceptance clears it: it needs a person's review outside Socair.

VERIFICATION

Signing: DSSE Ed25519 over an in-toto Statement v1

Signer key: 6230d2225405c855da4cfba73cad678651dd6f0221e2739fc9bd3901ffd35e38

Document hash: 4413dd4850dd5b1b4c96b674895a072a9abeb9d6947009f724178719c1eb04a8

A printed report is a claim, not the proof. Check the attestation: socair verify <attestation.dsse.json> --trusted <key.pub>

--artifact <file or directory>

Artifact SHA256: aae760a874f44240faefc3f21b3907b7379822ef6fb1c38baa64b336c25538f2

FILES

The artifact SHA256 is the digest of this manifest: every file's hash, size, and path (socair.modeldir/v1).

.gitattributes · other · 1519 bytes · 11ad7efa24975ee4b0c3c3a38ed18737f0658a5f75a0a96787b576a78a023361
README.md · other · 1184 bytes · a4cd79d07103c8b99e1eb0a0b2a66d82086f981e82cf9524943db9de547aaa81
config.json · config · 1637 bytes · 8d0c3b7a0275c5ffa96646f19d0cc745373870775beb590a22e7f6342445931d
configuration_chatglm.py · code · 2332 bytes · 351ba435758670b27c160b7657c7ab1d32053c4e8c0606a5a9b8f9908249dec8
generation_config.json · config · 111 bytes · dd87b7043c1a26d71a3d42bcd2c89c95f189861b6dd50d54b882b11859f5a527
model.safetensors · weights · 77305592 bytes · feebf1e929365645e986ae056b8fc1ab8c56f96f44d837f4d2cdd002b4900a44
modeling_chatglm.py · code · 56383 bytes · f937ce86e1906317003f043227b5a341c4870dc230c1324ca251facf2085d605
quantization.py · code · 14692 bytes · 7a690346ab3866ebdc361fb5d7389ecf6e3e8fe62f50868d28a5bcbff52ecd20
special_tokens_map.json · tokenizer · 3 bytes · ca3d163bab055381827226140568f3bef7eaac187cebd76878e0b63e9e442356
tokenization_chatglm.py · code · 10901 bytes · 8c81dd8b268f4bc9f9730b2cd5d9cf39d52809acb5c7d75e579c57358f7c5ddd
tokenizer.model · tokenizer · 1018370 bytes · e7dc4c393423b76e4373e5157ddc34803a0189ba96b21ddbb40269d31468a6f2
tokenizer_config.json · tokenizer · 458 bytes · 88119a30dc2458052f642fbcd86c8a41acd7977ced52dae4e4e9ae9ca3415356

socair.ai examples · No absence guarantee. See the bounded statement and out-of-scope ceiling.