

Model Assurance Attestation

Issued 2026-10-15T03:08:50Z · Template 0.1

hf-internal-testing/tiny-random-bert

69b4b82f6d6c223aa4fe2a958db6942de819c80a878829beeffa116248d152f7

model directory · document SOCAIR-20261015-69b4b82f · issuer socair.ai examples

no license stated

Tier 1 (static)

Promotion: authorized with conditions

Accepted, not tested: Format and structure. Accepted by Example reviewer on 2026-10-15T03:08:50Z (signed acceptance, until 2027-01-13T03:08:50Z).

5 pass 0 fail 0 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	NOT_TESTED	weights are pickle checkpoints; their structure is read by the pickle row
----------------------	--	------------	---

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	repo: 7 files, 0 scripts inventoried (scripts are expected, not a finding). 1 archive(s) read entry by entry by another check, with no entry left unread: pytorch_model.bin (Pickle opcode scan: archive/data.pkl conforms to grammar socair-wo/1; 100 storage record(s) of exactly numel x itemsize bytes; torch.save metadata version)
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/hf-internal-testing/tiny-random-bert; revision f171d7baecaf37b5da5a3616d8833b9969753535; commit f171d7baecaf37b5da5a3616d8833b9969753535; a selection of the repo's files at that commit (--exclude model.safetensors --exclude tf_model.h5 --exclude onnx/), 3 left out: model.safetensors, onnx/model.onnx, tf_model.h5; via airlock pull; publisher signature not established; manifest provenance.json (sha256 f1b30838307306889ee2d46cac1a84fb795681c6c1a397da28463409fbe18621)

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_REFERENCE), 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	PASS	tokenizer.json (sha256 dd65e2c292bcc86c, WordPiece model): 1124 tokens, 5 added (5 special); special-token ids, added-token tables, declared special tokens, and template tokens agree; no token, normalizer, or template injects text. No canonical tokenizer reference was configured, so its tokens were not compared.
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	PASS	no loader is told to import code: no auto_map, custom_pipelines, sentence-transformers module outside sentence_transformers, or diffusers custom component in the 4 JSON config(s) read, and no custom_generate/generate.py, handler.py, pipeline.py, or code/inference.py

Pickle opcode scan	Imports in pickle-based model files that reach code execution, and pickles that build more than tensors and plain containers PASS means: Every pickle conforms to grammar socair-wo/1: protocol 2 or 3 as CPython writes it, building only tensors and plain containers through calls whose arguments were checked, and every storage it references matches a record of exactly its size. A NumPy array with no object dtype holds no pickle, and its data matches its shape. Not covered: a loader bug, a reader that parses the bytes differently, code carried as data, or a tampered Python environment. Addresses AML.T0011.000, AML.T0018.002, AML.T0010.003, LLM03	PASS	1 of 1 file(s) PASS (pytorch_model.bin); e.g. pytorch_model.bin: zip archive: 1 pickle(s) conform to grammar socair-wo/1 (protocol 2); they build only tensors and plain containers: 101 tensor(s) (float32 x100, int64 x1), shapes [1 512], [1124 32], [512 32], and 98 more. torch zip: 100 storage record(s), each exactly numel x itemsize bytes, and no record torch.load does not read. Not covered: a bug in a loader, a reader that parses these bytes differently from CPython, code carried as data in a string for a later stage, and a tampered Python environment (a patched module).
--------------------	---	------	--

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 that returned PASS found no indicators within their stated scope. 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: authorized with conditions · Tier 1 only

Authorized with conditions: Format and structure is NOT_TESTED and were accepted by Example reviewer (signed acceptance, acceptor key 0085cfcd39d4838f) until 2027-01-13T03:08:50Z.

VERIFICATION

Signing: DSSE Ed25519 over an in-toto Statement v1

Signer key: 6230d2225405c855da4cfba73cad678651dd6f0221e2739fc9bd3901ffd35e38

Document hash: 8ba702045f52f8cc114baf8563cb9d0aeeb4e4182c74cba0c07e9fd13992eb6c

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: 69b4b82f6d6c223aa4fe2a958db6942de819c80a878829beeffa116248d152f7

FILES

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

.gitattributes · other · 737 bytes · 8072d6759ec86b43715d0f8d0b591e411f34d342e34ce6ab422ad205f4bbc114

config.json · config · 548 bytes · 54f9f0001ec62b6a53072a7c9d1b630e346ae51fca14a5799e54f0d918281872

pytorch_model.bin · weights · 540217 bytes · 9922e8996d0c7e24c7f4e7a5d9c5b7303549f4ee94de0f1138b103014b51be13

special_tokens_map.json · tokenizer · 112 bytes · 303df45a03609e4ead04bc3dc1536d0ab19b5358db685b6f3da123d05ec200e3

tokenizer.json · tokenizer · 12930 bytes · dd65e2c292bcc86c93a547f53117b419711310c1467a90da2aac6ec6d5ee1c6b

tokenizer_config.json · tokenizer · 321 bytes · 9a8ed9b01c8a56b555dc31cd526ba9e488cfc16ee5a101b3ce64af81d34f3d

vocab.txt · tokenizer · 4679 bytes · 130f1804e2dd52ff08aee3df4c216167f86e7ee4e2fed48bbe0537d61e8dd117

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