

Model Assurance Attestation

Issued 2026-10-15T03:08:46Z · Template 0.1

Qwen/Qwen3-0.6B

fb2e51d2156914d13091c694e8ac226bedc3022c587f31f72ca8831304fb199
model directory · document SOCAIR-20261015-fb2e51d · issuer socair.ai examples
license Apache License 2.0 (apache-2.0) · declared base Qwen/Qwen3-0.6B-Base

Tier 1 (static)

Promotion: authorized

7 pass 0 fail 0 lead 0 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: 311 tensors tile 1503264768 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: 10 files, 0 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/Qwen/Qwen3-0.6B; revision c1899de289a04d12100db370d81485cdf75e47ca; commit c1899de289a04d12100db370d81485cdf75e47ca; via airlock pull; publisher signature not established; manifest provenance.json (sha256 922518413c4ee08b44acce0fff52d24722416e86579ed6e5e37b42facf7f42f3)

Known-bad hash match	Match against the known-bad artifact denylist PASS means: This exact hash is not on the configured denylist. Any changed byte gives a new hash, so it catches only known files. Addresses AML.T0115.001, AML.T0010.003, LLM03	PASS	11 of 11 file(s) PASS; e.g. (directory manifest): no entry matched among 1 known-bad hashes (denylist denylist.txt (sha256 19db9513b112f5c7383ad54f8eb20183b26a02c46b3daf7c2429a4b25897df64))
Chat template	Code reach, hidden or obfuscated text, override or content-triggered instructions, and the text the rendered template adds to the prompt PASS means: The template parsed and reaches no Python internals or process execution, and carries no override or concealment language, URL, hidden or obfuscated text, or condition on message content other than a standard content-part type such as text or image. Instructions written as ordinary guidance, or added for a content-part type, are not judged: the notes list the text a template adds for a content-part type, and, when the template rendered, the text it adds to the prompt and whether it matches a reviewed template. Addresses AML.T0018.003, AML.T0051, AML.T0011.000, LLM01, LLM03	PASS	template parsed; no code-execution reach, content-conditional injection, hidden or obfuscated text, or instruction-language lead found. Structural heuristic; not a proof of absence. Rendered on 6 probes/v1 probe conversation(s) with Socair's own Jinja evaluator, modelling transformers' sandbox settings. This template adds to the prompt: "# Tools\n\nYou may call one or more functions to assist with the user query.\n\nYou are provided with function signatures wi..." (2 of 6); "</tools>\n\nFor each function call, return a json object with function name and arguments within" (2 of 6); "XML tags:" (2 of 6); "{\n\"name\":<function-name>, \"arguments\":<args-json-object>}" (2 of 6); "{\n\"name\": \"\" (1 of 6); "\", \"arguments\": \"\" (1 of 6). Listed for review, not judged. No reviewed template text was available, so it was not compared.

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 aeb13307a71acd8f, BPE model): 151669 tokens, 26 added (14 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 5 JSON config(s) read, and no custom_generate/generate.py, handler.py, pipeline.py, or code/inference.py

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

Node classes not tested:

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

PROMOTION AUTHORIZATION

Clean-store promotion: authorized · Tier 1 only

Every Tier 1 check PASSED.

VERIFICATION

Signing: DSSE Ed25519 over an in-toto Statement v1

Signer key: 6230d2225405c855da4cfba73cad678651dd6f0221e2739fc9bd3901ffd35e38

Document hash: 46b6f0f839d0faf56278b9eb50b2d56155a75b2e3dd2315ceeb8546c6244c10d

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: fbb2e51d2156914d13091c694e8ac226bedc3022c587f31f72ca8831304fb199

LICENSE STATEMENTS

What the artifact says about its license, and where. A license is identified, never reviewed: this is not legal advice.

LICENSE: "Apache License" · apache-2.0

model card license: "apache-2.0" · apache-2.0

model card license_link: "https://huggingface.co/Qwen/Qwen3-0.6B/blob/main/LICENSE" · not identified (a link; the scan does not follow links)

FILES

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

.gitattributes · other · 1570 bytes · 34448b82c17d60fec9b65b1f093c115ddbaadc04beb1b0140b6bfed2e012a930

LICENSE · other · 11343 bytes · 832dd9e00a68dd83b3c3fb9f5588dad7dcf337a0db50f7d9483f310cd292e92e

README.md · other · 13965 bytes · 1ab64a26fcb3b461423b89a433a8c858f1bf8d4086f979cbb3ff878d47cf20e9

config.json · config · 726 bytes · 660db3b73d788119c04535e48cf9be5f55bc3100841a718637ae695b442f27dd

generation_config.json · config · 239 bytes · 2325da0f15bb848e018c5ae071b7943332e9f871d6b60e2ed22ca97d4cb993d2

merges.txt · tokenizer · 1671853 bytes · 8831e4f1a044471340f7c0a83d7bd71306a5b867e95fd870f74d0c5308a904d5

model.safetensors · weights · 1503300328 bytes · f47f71177f32bcd101b7573ec9171e6a57f4f4d31148d38e382306f42996874b

tokenizer.json · tokenizer · 11422654 bytes · aeb13307a71acd8fe81861d94ad54ab689df773318809eed3cbe794b4492dae4

tokenizer_config.json · tokenizer · 9732 bytes · d5d09f07b48c3086c508b30d1c9114bd1189145b74e982a265350c923acd8101

vocab.json · tokenizer · 2776833 bytes · ca10d7e9fb3ed18575dd1e277a2579c16d108e32f27439684afa0e10b1440910

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