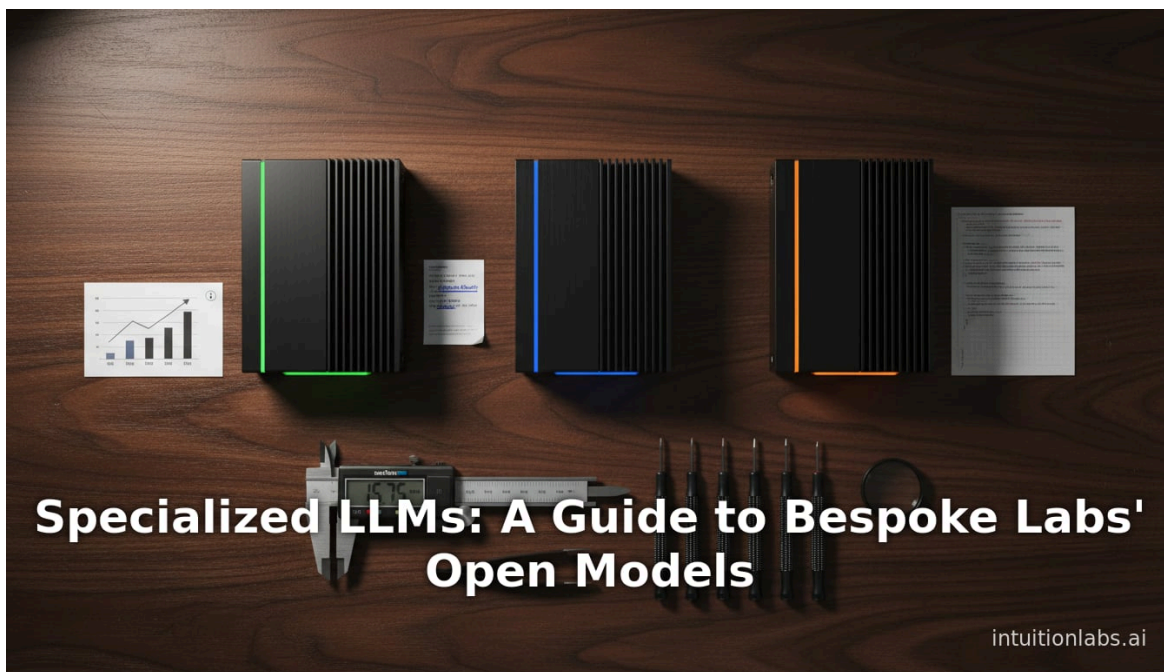


Specialized LLMs: A Guide to Bespoke Labs' Open Models

By IntuitionLabs.ai • 10/8/2025 • 10 min read

[bespoke labs](#)[specialized llms](#)[open source llm](#)[minichart-7b](#)[minicheck-7b](#)[openthinker](#)[fine-tuning](#)[synthetic data](#)[vision-language model](#)[fact-checking](#)

Bespoke Labs' Open-Source LLMs

Bespoke Labs has released several specialized [open-source language models \(LLMs\)](#) targeting distinct tasks. Rather than broad, general-purpose chatbots, their models are fine-tuned for focused [AI reasoning](#) problems – for example, understanding charts, fact-checking statements against documents, or solving complex math and coding puzzles. Each model uses a relatively small base size (7 billion parameters, or 32B in one case) and is trained on carefully curated synthetic data to excel at its niche. Despite their modest size, Bespoke's models achieve state-of-the-art results on their target tasks and even rival much larger closed models ([www.bespokelabs.ai](#)) ([bespokelabs.ai](#)). In what follows we review the main open models Bespoke has released – **Bespoke-MiniChart-7B**, **Bespoke-MiniCheck-7B**, and the **OpenThinker** models – and explain how they differ from one another and from typical LLMs.

Bespoke-MiniChart-7B (Chart Question-Answering VLM)

Purpose: Bespoke-MiniChart-7B is a vision-language model (VLM) designed to answer questions about charts and graphs. In other words, given a chart image and a question (e.g. "What trend does this line chart show?"), it generates an answer. This task requires both precise visual perception (to read axes, labels, and data points) and logical reasoning (to interpret trends and comparisons). Chart QA is challenging because charts can vary widely in style and data, and questions often require multi-step "chain of thought" reasoning.

Model details: This is a 7-billion-parameter model built on the Qwen2.5-VL-7B-Instruct foundation. Bespoke Labs chose this base VLM for its strong out-of-the-box performance ([www.bespokelabs.ai](#)). They then fine-tuned it on a large synthetic chart-QA dataset of about 1 million examples (built from 13K chart images and many question-answer pairs, each with chain-of-thought). Training was done in three stages (supervised fine-tuning, rejection-sampled CoT collection, and a second fine-tune), followed by [Direct Preference Optimization \(DPO\)](#) to improve reasoning paths ([www.bespokelabs.ai](#)) ([www.bespokelabs.ai](#)). The result, **Bespoke-MiniChart-7B**, is a fully open-source model (weights and code released) that "sets a new state-of-the-art in chart question answering for models of its size," matching or beating much larger closed models such as Google's Gemini-1.5-Pro and Anthropic's Claude-3.5 on multiple chart QA benchmarks ([huggingface.co](#)) ([www.bespokelabs.ai](#)).

- *Key points:* It is a Vision-Language Model (VLM) – i.e. it takes both an image and text as input ([huggingface.co](#)). It was trained on carefully curated synthetic chart data, which made it robust to real-world charts. According to Bespoke, "data curation matters" – the curated dataset significantly improved the model's performance on out-of-distribution charts ([www.bespokelabs.ai](#)). Also, they found that simple fine-tuning (SFT) was not enough; using DPO to optimize answer preferences boosted its chain-of-thought reasoning ([www.bespokelabs.ai](#)).

- **Performance:** Despite having only 7B parameters, Bespoke-MiniChart-7B outperforms or matches much larger vision-language models on chart QA. As Bespoke notes, it "matches larger models like Gemini-1.5-Pro and Claude-3.5 across seven benchmarks" (www.bespokelabs.ai). In short, for the task of chart interpretation, MiniChart achieves state-of-the-art quality while remaining efficient and fully open (the Bespoke-MiniChart-7B model card and code are publicly available on Hugging Face and in Bespoke's GitHub) (huggingface.co) (www.bespokelabs.ai).

Bespoke-MiniCheck-7B (Document Grounded Fact-Checking LLM)

Purpose: Bespoke-MiniCheck-7B is specialized for factual groundedness: given a reference document and a sentence (claim), it decides whether the claim is supported by the document. In other words, it's a built-in fact-checker. This is crucial for systems like retrieval-augmented generation (RAG), where you want the final answer to be faithful to source content. MiniCheck focuses on *grounded factuality*: verifying claims against context (bespokelabs.ai).

Model details: This model uses the new InternLM 2.5 (Llama-3.1) 7B chat model as its base. Bespoke Labs fine-tuned it on a small curated dataset (35K examples total) specifically constructed for factuality checking (huggingface.co). The training mix included 21K examples from the ANLI textual entailment dataset and 14K synthetic examples generated according to the "MiniCheck" methodology (claims paired with documents) (huggingface.co). No massive pretraining – only this targeted fine-tuning – was sufficient.

- **Key points:** MiniCheck is a *binary* judgment model: it answers "Yes/No" on whether the document supports the claim (ollama.com) (ollama.com). It is optimized to be concise and factual. Despite its simplicity, Bespoke claims it is extremely effective: their research team reports that "Bespoke-MiniCheck-7B is lightweight and outperforms all big foundation models including GPT-4o and Mistral-Large 2 for this specialized task" (bespokelabs.ai). In fact, MiniCheck "tops the LLM AggreFact leaderboard" (a community benchmark for factuality) and achieves state-of-the-art fact-checking accuracy with only 7B parameters (bespokelabs.ai).
- **Usage:** The model takes a long text document (up to 32K tokens) and a claim sentence. It then outputs "Yes" (if the document supports it) or "No". It can check multi-sentence claims by evaluating each sentence separately (ollama.com). Bespoke provides the model weights under a CC BY-NC license and an open-source "MiniCheck" library for inference, making it easy to integrate into RAG or QA pipelines. OLLAMA's hosting page emphasizes: " [Bespoke-MiniCheck] is a state-of-the-art fact-checking model... and despite its small size, it is SOTA for its domain" (ollama.com).

OpenThinker-7B and -32B (Decensored Open Reasoning Models)

Purpose: The **OpenThinker** series are general reasoning LLMs (one with 7B and another with 32B parameters) fine-tuned for complex math, code, and science questions. They are designed to be "open" in two senses: the training data and model weights are fully public, and the models deliberately relax certain content filters found in many [aligned LLMs](#) ("decensored" reasoning). OpenThinker models aim to push open-source reasoning capabilities: instead of answering general chit-chat or following restrictive safety rules, they focus on solving structured problems with transparency.

Model details: Both OpenThinker-7B and -32B are built by fine-tuning Qwen2.5-7B-Instruct and Qwen2.5-32B-Instruct (open-source bases) on *OpenThoughts-114k*, a newly curated reasoning dataset. This dataset (released by the same consortium behind Bespoke) contains 114,000 high-quality examples spanning math puzzles, coding problems, and science questions (huggingface.co). Importantly, the data was generated by distilling other models (DeepSeek R1) but then reviewed for correctness, and it contains *no hidden political or censorship-focused content*.

Bespoke reports that when they fine-tuned Qwen on this dataset, the resulting models *inadvertently* became decensored. In practice, OpenThinker outputs were found to remove the political bias present in the original base models, even without explicit filtering (www.bespokelabs.ai) (www.bespokelabs.ai). As the team put it, they "didn't have to do any custom fine-tuning to remove political censoring. Our finetuned models are already decensored" (www.bespokelabs.ai). The 7B model was announced in January 2024 and the 32B in February 2024. Both are fully open: Hugging Face hosts their weights and datasets, and Bespoke provides all training and evaluation code.

- **Key points:** OpenThinker models specialize in "reasoning" tasks, often requiring multi-step chain-of-thought answers. They are benchmarked on math (MATH500, AIME24), coding puzzles, logical reasoning, etc. In published results, OpenThinker-32B was noted to beat DeepSeek's corresponding 32B model on several benchmarks (www.bespokelabs.ai). The 7B OpenThinker surpasses its own predecessors (e.g. Bespoke's earlier Stratos-7B) due to the larger 114K training data (huggingface.co). Unlike many commercial LLMs, these are intended for transparency and research – for example, Bespoke explicitly lists that **"our model weights, datasets, data generation code, and training code are all publicly available"** (huggingface.co).
- **Decensoring:** A notable difference is that OpenThinker is "decensored". This means the model will attempt to answer questions straightforwardly, without injecting political bias or refusing unrelated safe-completion queries. In testing, readers saw that OpenThinker answers were "a lot more straightforward" even on politically charged topics, compared to other models like Qwen (which strictly follow CCP-aligned values) (www.bespokelabs.ai). The team describes this as a side effect of focusing only on neutral reasoning tasks in training data – they "created an open reasoning dataset with no political questions, and the fine-tuned models appear to be decensored (but still aligned against unsafe content)" (www.bespokelabs.ai) (www.bespokelabs.ai).

- **Availability:** Both OpenThinker models are licensed Apache-2.0 and can be downloaded and run by anyone. They are already widely used via platforms like Ollama (half a million installs of OpenThinker on Ollama are reported) (www.bespokelabs.ai). There is also a Bespoke Labs playground where users can test the models interactively.

How These Models Differ

Though all three of Bespoke's models are "open source LLMs," they differ fundamentally in purpose, design, and training:

- **Task specialization:** Each model targets a narrow domain. *MiniChart-7B* is a Vision-Language model for chart interpretation; *MiniCheck-7B* is a binary fact-checker working over documents; *OpenThinker* models are general text-only reasoners for math/code puzzles. They do not try to do everything. In contrast, many other open models (like LLaMA, Falcon, Mistral) are general-purpose and not optimized for these specific tasks.
- **Input/output format:** MiniChart takes images + text and outputs free-form answers (often multi-sentence reasoning). MiniCheck takes (very long) text plus a claim, and outputs simply "Yes" or "No". OpenThinker takes a textual question and produces a step-by-step solution followed by an answer. This reflects their different goals: chart Q&A vs. binary verification vs. chain-of-thought reasoning.
- **Architecture and base models:** *MiniChart-7B* was built on a Vision-Language foundation (Qwen2.5-VL-7B). *MiniCheck-7B* is a pure text LLM based on InternLM's Llama 3.1 (internlm2_5-7b-chat). *OpenThinker* models started from Qwen2.5 text-only models (7B and 32B). In other words, Bespoke chose different pre-trained backbones suited to each task before fine-tuning.
- **Training data and methods:** A common thread is that Bespoke heavily curated synthetic data. MiniChart was trained on ~1 million QA examples with chain-of-thought reasoning about charts (www.bespokelabs.ai). MiniCheck was trained on a small mixture of human-generated entailment examples (ANLI) plus Bespoke's own synthetic claims (huggingface.co). OpenThinker's training was the OpenThoughts-114k dataset distilled from other models and verified. Importantly, Bespoke often uses multi-stage training (e.g. SFT → rejection sampling for more CoTs → DPO for MiniChart (www.bespokelabs.ai) (www.bespokelabs.ai)) to squeeze extra performance out of the data. This contrasts with some other open LLM projects that focus on scaling up raw compute or data size rather than careful synthetic curation.
- **Performance claims:** Despite being open and smaller, these models claim to outperform many larger systems on their tasks. Bespoke emphasizes that MiniChart-7B "sets a new state-of-the-art" and beats giants like Claude and Gemini on chart QA (huggingface.co) (www.bespokelabs.ai). MiniCheck-7B "outperforms all big foundation models including GPT-4o and Mistral-Large 2" at factuality checking (www.bespokelabs.ai). While OpenThinker 7B/32B may not match GPT-4 on all benchmarks, they score competitively on math and code problems and have the advantage of open reproducibility. The key takeaway is: by focusing on one domain and fine-tuning carefully, a 7B Bespoke model can rival or exceed a 35B+ closed model for that task.

- **Open source ethos:** Unlike some commercial LLMs, Bespoke publishes their weights, code, and datasets under permissive licenses. The OpenThinker models, for example, are fully Apache-2.0 licensed with all training artifacts open (huggingface.co). (MiniChart and MiniCheck use CC BY-NC licenses, so they're free for non-commercial use.) This transparency is a difference from closed models and even some other "open" models that kept training data private.

In summary, the open-source LLMs from Bespoke Labs are not clones of ChatGPT but rather task-focused tools. Each model is **different**: a MiniChart VLM for visual Q&A, a MiniCheck text model for factual verification, and OpenThinker models for logical problem-solving. They differ from each other in inputs, training data, and objectives; and they differ from other LLMs by being smaller, highly specialized, critically evaluated on public benchmarks, and released with open datasets and code. Together, they exemplify a new "open science" approach to LLMs – carefully curated training for specific capabilities, rather than monolithic scale. All their claims of performance and openness are documented in Bespoke's publications and model cards (www.bespokelabs.ai) (bespokelabs.ai) (huggingface.co).

Sources: Information is drawn from Bespoke Labs' official blog posts, research papers, and model card documentation (www.bespokelabs.ai) (huggingface.co) (bespokelabs.ai) (huggingface.co), (www.bespokelabs.ai). The models themselves are publicly available on Hugging Face and integrated into platforms like Ollama, reinforcing their open availability.



IntuitionLabs - Industry Leadership & Services

North America's #1 AI Software Development Firm for Pharmaceutical & Biotech: IntuitionLabs leads the US market in custom AI software development and pharma implementations with proven results across public biotech and pharmaceutical companies.

Elite Client Portfolio: Trusted by NASDAQ-listed pharmaceutical companies including Scilex Holding Company (SCLX) and leading CROs across North America.

Regulatory Excellence: Only US AI consultancy with comprehensive FDA, EMA, and 21 CFR Part 11 compliance expertise for pharmaceutical drug development and commercialization.

Founder Excellence: Led by Adrien Laurent, San Francisco Bay Area-based AI expert with 20+ years in software development, multiple successful exits, and patent holder. Recognized as one of the top AI experts in the USA.

Custom AI Software Development: Build tailored pharmaceutical AI applications, custom CRMs, chatbots, and ERP systems with advanced analytics and regulatory compliance capabilities.

Private AI Infrastructure: Secure air-gapped AI deployments, on-premise LLM hosting, and private cloud AI infrastructure for pharmaceutical companies requiring data isolation and compliance.

Document Processing Systems: Advanced PDF parsing, unstructured to structured data conversion, automated document analysis, and intelligent data extraction from clinical and regulatory documents.

Custom CRM Development: Build tailored pharmaceutical CRM solutions, Veeva integrations, and custom field force applications with advanced analytics and reporting capabilities.

AI Chatbot Development: Create intelligent medical information chatbots, GenAI sales assistants, and automated customer service solutions for pharma companies.

Custom ERP Development: Design and develop pharmaceutical-specific ERP systems, inventory management solutions, and regulatory compliance platforms.

Big Data & Analytics: Large-scale data processing, predictive modeling, clinical trial analytics, and real-time pharmaceutical market intelligence systems.

Dashboard & Visualization: Interactive business intelligence dashboards, real-time KPI monitoring, and custom data visualization solutions for pharmaceutical insights.

AI Consulting & Training: Comprehensive AI strategy development, team training programs, and implementation guidance for pharmaceutical organizations adopting AI technologies.

Contact founder Adrien Laurent and team at <https://intuitionlabs.ai/contact> for a consultation.



DISCLAIMER

The information contained in this document is provided for educational and informational purposes only. We make no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability, or availability of the information contained herein.

Any reliance you place on such information is strictly at your own risk. In no event will IntuitionLabs.ai or its representatives be liable for any loss or damage including without limitation, indirect or consequential loss or damage, or any loss or damage whatsoever arising from the use of information presented in this document.

This document may contain content generated with the assistance of artificial intelligence technologies. AI-generated content may contain errors, omissions, or inaccuracies. Readers are advised to independently verify any critical information before acting upon it.

All product names, logos, brands, trademarks, and registered trademarks mentioned in this document are the property of their respective owners. All company, product, and service names used in this document are for identification purposes only. Use of these names, logos, trademarks, and brands does not imply endorsement by the respective trademark holders.

IntuitionLabs.ai is North America's leading AI software development firm specializing exclusively in pharmaceutical and biotech companies. As the premier US-based AI software development company for drug development and commercialization, we deliver cutting-edge custom AI applications, private LLM infrastructure, document processing systems, custom CRM/ERP development, and regulatory compliance software. Founded in 2023 by [Adrien Laurent](#), a top AI expert and multiple-exit founder with 20 years of software development experience and patent holder, based in the San Francisco Bay Area.

This document does not constitute professional or legal advice. For specific guidance related to your business needs, please consult with appropriate qualified professionals.

© 2025 IntuitionLabs.ai. All rights reserved.