

AMERICAS TECHNOLOGY: SOFTWARE

Nine Conversations in Silicon Valley last week: MSFT, SNOW, Security, Robotics

We had the opportunity to spend time with industry experts alongside Microsoft Build and the Snowflake Summit the week of June 1 in San Francisco.

Both companies in our view did several thoughtful things to set up for stock outperformance over the next year. Microsoft directly demonstrated how domain experience and customer-specific context can drive superior agentic outcomes; both with its MAI-Thinking-1 medium language model, and its introduction of Frontier Tuning for customers to train their own models. Snowflake published benchmarks of CoCo (Cortex Code) outperformance vs. 3P coding tools, and customers firmly pushed back when we tested our AI bear cases on new competition and value disintermediation. Although one inflection in consumption is already happening in 1H (and reflected in the stock), there are more levers to pull, including with a likely accelerated launch of Iceberg v4 with multimodal support in the coming months — a very tangible example of how Snowflake has accelerated product innovation. **We remain Buyers of both stocks and also include an update on Security and Robotics within.**

Our observations:

1) SLM > LLMs, and introducing MLMs: One of our key observations over the last 3 months is that an efficient frontier of intelligence is coming into focus: rather than tokenmaxxing and throwing every problem at a LLM, customers are now being more thoughtful about using more fine-grained (and thus cheaper and potentially more performant) tools for specific tasks. In response, vendors are providing more fine-grained tools that are in turn differentiated by being trained for specific use cases where they have proprietary domain experience. (Datadog was one of the first companies we saw illustrate this concept at their [analyst day](#) in February). This creates a spectrum of SLMs that can solve every day intelligence tasks (e.g., those that require an IQ of <100), while the frontier labs can continue to push the bleeding edge of intelligence problems (e.g., those that perhaps surpass the top decile of human intelligence). The most substantial parts of the MSFT and SNOW conferences this week were their new disclosures on the specific ways in which their models outperform the industry.

- For MSFT, benchmarking shows their MAI-Thinking-1 outperforming on specific STEM and coding problems, and being >10x more cost-efficient with no distillation from 3P models. MAI-Thinking-1 is a Medium Language Model (MLM):

Gabriela Borges, CFA
+1(212)902-7839 |
gabriela.borges@gs.com
Goldman Sachs & Co. LLC

Maura Hager
+1(212)9028724 | maura.hager@gs.com
Goldman Sachs & Co. LLC

Selina Zhang
+1(212)357-9979 |
selina.zhang@gs.com
Goldman Sachs & Co. LLC

Max Gamperl
+1(415)249-7311 |
max.gamperl@gs.com
Goldman Sachs & Co. LLC

Goldman Sachs does and seeks to do business with companies covered in its research reports. As a result, investors should be aware that the firm may have a conflict of interest that could affect the objectivity of this report. Investors should consider this report as only a single factor in making their investment decision. For Reg AC certification and other important disclosures, see the Disclosure Appendix, or go to www.gs.com/research/hedge.html. Analysts employed by non-US affiliates are not registered/qualified as research analysts with FINRA in the U.S.

it has a smaller inference footprint than the LLMs but can meet them selectively on performance. Investors have been questioning Microsoft’s strategy in AI for several quarters. Microsoft’s 1P model innovations and its detail around how enterprises can train MAI-T-1 on their own data crystallizes its strategy and also shows their progress as an independent AI company. Both milestones we think will help remove a key overhang from the stock.

- Snowflake disclosed benchmarks that showed CoCo outperforming other coding tools in the Snowflake environment, and our conversations on the show flow confirmed this outperformance in practice in customer environments.

Exhibit 1: MAI-Thinking-1 benchmarks comparably to other frontier reasoning models

MAI-Thinking-1 Metrics vs. Other Reasoning Models

Category	Benchmark	MAI-Thinking-1	Sonnet 4.6	Opus 4.6	GPT 5.4	Kimi K2.6	DeepSeek V3.2	DeepSeek V4	GLM-5.1
STEM	AIME 2025	97.0	95.6	99.8	-	-	93.1	-	-
	AIME 2026	94.5	-	-	-	96.4	-	-	95.3
	HMMT Feb 2026	84.9	-	-	-	92.7	-	95.2	82.6
	GPQA Diamond	84.2	89.9	91.3	92.8	90.5	82.4	90.1	86.2
	LCB v6	87.7	-	-	-	89.6	83.3	93.5	-
Agentic Coding	Terminal Bench 2.0	46.0	59.1	65.4	75.1	66.7	46.4	67.9	69
	SWE-Bench Verified	73.5	79.6	80.8	-	80.2	73.1	80.6	-
	SWE-Bench Pro	52.8	-	53.4	57.7	58.6	-	55.4	58.4

Source: Company data, Goldman Sachs Global Investment Research, Artificial Analysis

Exhibit 2: MAI-Code-1-Flash outperforms Claude Haiku 4.5 across all core coding benchmarks

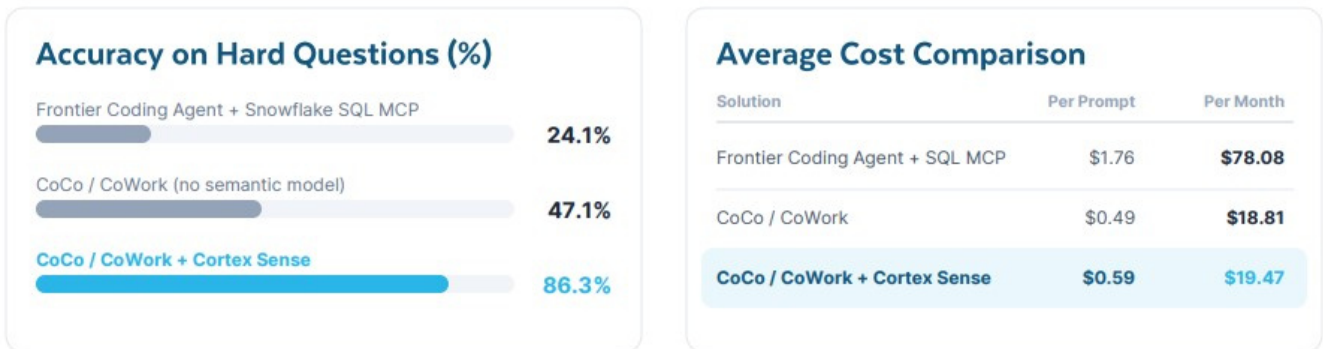
MAI-Code-1-Flash vs. Claude Haiku 4.5

Category	Benchmark	MAI-Code-1-Flash		Claude Haiku 4.5	
		Pass Rate	Average Token Usage (K)	Pass Rate	Average Token Usage (K)
Agentic Coding	SWE-Bench Verified	71.6	10.8	66.6	27.3
Diverse Agentic Coding	SWE-Bench Pro	51.2	28	35.2	29.8
Multilingual Coding	SWE-Bench Multilingual	65.5	15.3	62.7	17.2
Agentic Terminal Coding	Terminal Bench 2.0	54.8	21.6	41.6	25

Source: Company data, Goldman Sachs Global Investment Research, Artificial Analysis

Exhibit 3: Snowflake illustrates that context raises accuracy and lowers cost

Sample eval set



Source: Company data

2) On headless and the future of apps: We heard several supporting arguments for the future of app software being headless, but the debate lies as to whether app software companies can monetize in a headless world. The clearest examples of headless we have

seen so far are at Cloudflare (where their sales tools have been custom built on top of Salesforce) and at Anthropic (which detailed their sales stack at SaaStr with Claude built on top of Salesforce, Slack, Gong, Intercom, Ironclad, and others). We also heard that both Claude and Gemini Enterprise are making progress at becoming a de facto knowledge worker search interface (Google's experience with intuitive consumer-facing UI is now extending into the enterprise and being complemented by the raw power of Gemini). In this context, the UI sits on top of the AI control plane, which now becomes the platform layer in the IPA stack (infra, platform, apps), and simultaneously gives the data warehouse companies more right to win in applications. That said, industry feedback on Snowflake CoWork was mixed with some noting it does not have clear product market fit. Our best guess today is that there will be bifurcation between the most sophisticated and typical customer, as this seems to be an extension of the build vs buy debate. More sophisticated customers are likely to build more native apps (Cloudflare), less sophisticated customers are likely to simply use 3P agentic apps as an extension of their SaaS provider. The question for the underlying SoRs remains: if engagement via API access goes up, how do they monetize it? On the one hand, more usage of data in an SoR could increase its strategic value; on the other hand, most customers we spoke to noted that it would be challenging for SoR companies to monetize this usage.

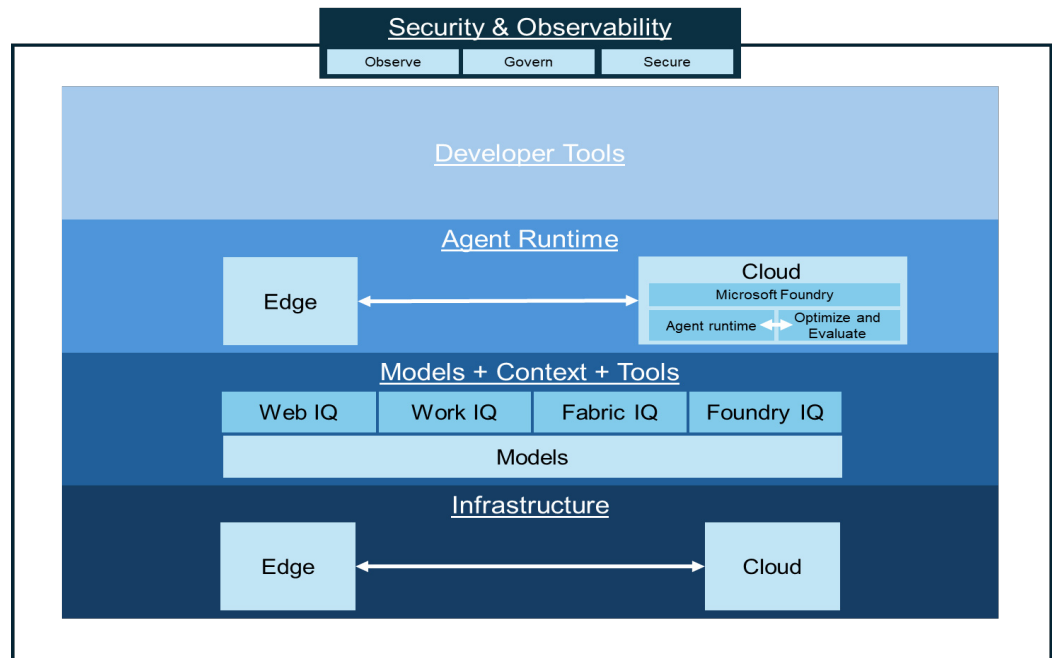
3) On Security: our industry conversations this week, including in callbacks with **CrowdStrike** and **Palo Alto Networks**, focused on the lag effect between infrastructure spending and security spending; first, in the cloud cycle 2015-2020; and now, in the AI enterprise cycle, 2025+. In the cloud cycle, we estimate about a 2-year lag between IaaS scaling and cloud security inflecting from <1% of infrastructure spend to >3% of spend. Today, in the AI cycle, we believe a similar 3-5% benchmark is appropriate (consistent with comments from Palo Alto Networks), and that the lag will likely be <2 years. If we consider year 1 of enterprise adoption as 2026 and triangulate our customer conversations, we expect to see AI product cycles begin to move the needle for WholeCo security growth in 2H26 and into 2027. Our conversations suggest the lag is tied specifically to when we see more meaningful agentic enterprise adoption, at which point agentic containers will become more permanent and less ephemeral, and consensus will start to emerge on the best tools for agentic runtime security.

4) On Applied AI: We had the opportunity to meet with Generalist AI, an applied AI robotics company that raised \$400mn on 6/4. (Recall we also hosted Physical Intelligence at our Private Company conference in April and our colleague Mark Delaney hosted a Physical AI Bus Trip in March). Generalist is focused on training a generalist robotics model, with its key differentiation being a) using visual images rather than sensors. The model observes an off-the-shelf commercial robot via cameras, converts the signal into tokens, and uses the AI world model to predict the next best movement for the robot arms in space; b) training a proprietary model on real data in the real world to map the jagged frontier of what's possible, rather than simulated data (which has been the traditional approach for rigid-body physics like the backbone of self driving), creating a compounding empirical moat; c) a focus on dexterity, scaling, tightly integrated feedback loops between research and deployment, and a talent-dense team. In our view, companies like Generalist illustrate that the technical breakthroughs in AI in recent years may have meaningful implications well beyond the enterprise software TAM that we typically focus on. This may create a new class of companies that span the best of two business models: the Software business model (where code is the cornerstone of

driving model performance for real world applications) and those business models with a physical moat (which may be more defensible in an era where the cost of code is low).

Microsoft

Exhibit 4: Microsoft’s AI stack



Source: Company data, Goldman Sachs Global Investment Research

5) Bridging the gap between AI potential and AI reality: Microsoft introduced Frontier Tuning for enterprises to train models using their own data within existing compliance guardrails. Unlike traditional fine tuning, Frontier Tuning leverages managed reinforcement learning in real-world environments, where MAI models continuously learn and improve based on real workflows and agent interactions, allowing outputs more optimized for each business’s operations. (We believe this product is competitive with Amazon’s Nova Foundry, which was similarly well received at Reinvent). We see potential for this to accelerate enterprise AI adoption across Microsoft’s installed base, particularly as organizations look to balance outcome with rising inference costs. By domain specific tuning, Frontier Tuning can deliver materially better cost efficiency and ROI for production workloads. For example, MAI model fine-tuned for Excel is comparable to GPT 5.4 on public benchmarks while being 10x more cost-efficient, while MAI-Thinking-1 fine-tuned for McKinsey’s enterprise standards was >10x more cost-efficient than GPT 5.5.

6) Proofpoints on custom silicon: MAIA 200 performed well in theory based on benchmarks released on 1/26 (which we translated into our inference cost curve, [here](#)). We now have evidence that Microsoft can also ramp production and delivery performance improvements in the datacenter, with Maia 200 now live in Iowa and Arizona and deploying internationally in late-2026. Microsoft highlighted a 1.4x performance/watt improvement when training MAI-Thinking-1 end to end on Maia 200

vs. NVIDIA GB200. A more diversified silicon footprint should improve unit economics, particularly in the context of November 2025 commentary from CTO Kevin Scott that in house silicon could represent >50% of AI infrastructure over time. With NVIDIA and AMD gross margins of mid-70% and low- to mid-50%, respectively, we believe that vertically integrating 50% of the silicon footprint could improve Microsoft gross margin on Azure AI revenue by as much as ~20 points, albeit partly offset by pricing downs.

7) Expanding the intelligence layer: Microsoft emphasized that context quality is the primary driver of token efficiency. With the introduction of Web IQ alongside Work IQ, Fabric IQ, and Foundry IQ in the intelligence layer, Microsoft provides a continuously updated understanding of each organization, combining external knowledge (Web IQ), business ontology (Fabric IQ), and organizational knowledge graph (Work IQ), which is the grounded through Foundry IQ for agent execution. The key bottleneck to scaling agents in enterprise environments is not model capability, but how effectively context is structured for models. **This, in our view, is the answer to the question on how domain experience can drive superior outcomes: Microsoft establishes a single source of truth specific to the organization by leveraging its 50+ years of enterprise entrenchment, which in turn drives better customer-specific outcomes.**

Snowflake

8) On the sustainability of Cortex Code (CoCo) as a driver for Snowflake's business: Snowflake raised its FY27 revenue guidance by 3% in 1Q EPS, citing CoCo as the single biggest driver, and CEO Sridhar Ramaswamy disclosed on LinkedIn that early adopters consume 11% more of the platform. When we polled customers on the sustainability of growth tied to CoCo, the best response we got was: well, what do you think the sustainability of coding tools is? We think that 1H26 could show one step function higher in Snowflake usage at existing customers as the new pace of workload migrations from legacy databases, ETL tools, and reporting tools picks up (we heard between 30% to 70% improvement in migration speed – consistent with the fixed cost migrations being priced at consultant Squadron Data here). The next step higher will likely be as Snowflake boosts usage within an organization: as management described, the 7,000 customers that already use CoCo count just one use case at just one user in that total. There is still a lot of work to be done to make the product more accessible and more widely adopted. On this point, industry experts pointed out that while leadership has set a strong vision, they also have to bring the rest of the sales organization with them, and this requires wholesale training and enablement away from the classic SaaS sales pitch towards the agentic sales.

9) We kicked the tires on two structural bear cases on Snowflake: The first is new “AI native” competition such as Clickhouse (and potentially the frontier models), and the second is that AI democratizes data analytics for knowledge workers such that Snowflake's core business user can achieve similar outcomes without needing a dedicated data warehouse platform to structure and govern the data. We were surprised at just how few participants knew who Clickhouse was, suggesting that it may still be early in direct competition. On the second bear case, customers instead universally argued that democratization of data analytics makes Snowflake's platform more valuable: it has a very strong brand for data governance and the path of least resistance is to use more of the Snowflake platform. At the same time, both Amazon and Microsoft

seem to be somewhat agnostic as to whether customers choose their own data warehousing products vs. Snowflake, as Snowflake pulls through underlying compute and storage regardless, which means competition from hyperscalers has taken a step back (the exception is Google, which is still very competitive with BitQuery, particularly at CPG customers).

- We think Snowflake detailed a nice balance between communicating their strategy and their ability to execute, while acknowledging the meaningful pace of technology evolution still to come. One of our favorite comments from the Snowflake analyst session was that Software is changing so rapidly that people should not be in the business of making long-term predictions about what is possible and what is not. Instead, it was clear to industry participants that Snowflake's pace of innovation has picked up in the last 18 months, which will a) likely lead to an accelerated launch of Iceberg v4 with multimodal support e.g. for Lance; and b) help them to further entrench themselves for use cases where they have a right to win.

Valuation & Key risks

Microsoft (MSFT): We maintain our 12-month price target of \$610 (unchanged), based on a 28x P/E (unchanged) multiple to Microsoft's SNTM adjusted net income. Key downside risks include: longer ramp for internal silicon which can limit market share gains or gross margin expansion, greater investments in projects outside of expectations (e.g., non-Azure), key leadership changes, and more meaningful shift to custom software that could negatively impact its applications business.

Snowflake (SNOW): We increase our 12-month price target to \$300 (from \$278) based on 60x (from 55x on increased GAAP profitability - management introduced a GAAP profitability target by 4Q28, vs. -\$96mn/- \$177mn for GS/Street previously) EV/FCF multiple to Snowflake's SNTM (Q5-Q8) uFCF. Key downside risks include: 1) Iceberg cannibalizing Snowflake's storage revenue, 2) Increased competition from CSPs and Databricks, and 3) Adverse changes in the IT spending environment or further optimization cycles in cloud spending.

