Lesertest: Samsung 9100 Pro – Der Speicher-Turbo für lokale KI und Virtualisierung?

Gen 5 vs. Gen 4 vs. SATA: Ein Praxis-Check jenseits von CrystalDiskMark

Inhalt

1. Einleitung und Motivation

Speicherplatz ist heutzutage günstig, doch Speicher-*Geschwindigkeit* bleibt ein Luxusgut. Wer seinen PC nur für Office-Anwendungen oder Spiele nutzt, spürt zwischen einer guten PCIe 3.0 SSD und einem modernen PCIe 5.0 Modell oft kaum einen Unterschied. Doch wie sieht es aus, wenn der Rechner zur Workstation wird?

Als jemand, der sich gerne mit **lokaler Künstlicher Intelligenz (LLMs, Stable Diffusion)**, **Software-Entwicklung** und **Virtualisierung** beschäftigt, stelle ich mir eine entscheidende Frage:

"Lohnt sich der Aufpreis für die theoretische Bandbreite von PCIe 5.0 in der Praxis wirklich, oder wartet meine Grafikkarte am Ende doch nur auf die CPU?"

Synthetische Benchmarks wie CrystalDiskMark liefern zwar beeindruckende Zahlen, sagen aber wenig über das Verhalten beim Laden eines 8-Milliarden-Parameter-Sprachmodells oder beim Starten eines komplexen Docker-Containers aus. Mit diesem Lesertest der **Samsung 9100 Pro (2 TB)** möchte ich genau diese Lücke schließen. Mein Ziel ist es, nicht nur Balken zu vergleichen, sondern herauszufinden, ob und wo eine High-End-SSD im Creator-Alltag echte Lebenszeit spart.

2. Das Testfeld: David gegen Goliath

Um die Leistung der Samsung 9100 Pro fair einzuordnen, tritt sie nicht im luftleeren Raum an. Sie muss sich gegen eine breite Palette an Speichertechnologien behaupten, die stellvertretend für typische Aufrüst-Szenarien stehen:

- **Der Herausforderer (Gen 5):** Samsung 9100 Pro (2 TB) – *Das Testmuster*
- **Die Referenz (Gen 5):** Crucial T705 (1 TB) – *Der aktuelle "Speed-King"*
- **Der Vernunft-Standard (Gen 4):** Samsung 990 Pro (2 TB) – *Beliebt und bewährt*
- **Die Veteranen (SATA):** Crucial BX100 (250 GB) & OCZ Vertex 3 (120 GB) – *Ältere Modelle mit DRAM-Cache*
- **SATA Entry:** SanDisk Plus (120 GB) – *Günstiges Modell ohne DRAM-Cache*

3. Methodik: Reproduzierbarkeit statt Zufall

Wer SSDs unter Windows testet, kämpft gegen einen unsichtbaren Gegner: Das Betriebssystem selbst. Windows nutzt freien Arbeitsspeicher aggressiv als Datei-Cache (Standby List). Startet man eine Anwendung oder lädt ein KI-Modell zum zweiten Mal, kommen die Daten oft gar nicht mehr von der SSD, sondern direkt aus dem RAM. Das

Ergebnis wären traumhafte, aber völlig unrealistische Messwerte, die nichts mit der Leistung des Laufwerks zu tun haben.

Um echte "Real-World-Performance" zu messen, habe ich mich nicht auf manuelle Stoppuhren verlassen, sondern ein vollautomatisiertes Benchmark-Framework (PowerShell & Python) entwickelt.

3.1. **Der "Cold Start" Enforcer**

Das Herzstück meines Testskripts (run_all_benchmarks.ps1) ist die Kontrolle über den Windows-Cache. Vor jedem kritischen Messdurchlauf (gekennzeichnet als "Cold Run") wird das Tool EmptyStandbyList.exe ausgeführt.

- Funktion: Es zwingt Windows, den Datei-Cache im RAM sofort zu leeren.
- Effekt: Die SSD muss jedes Byte physisch neu lesen. Nur so lässt sich unterscheiden, ob eine SSD Daten mit 500 MB/s (SATA) oder 10.000 MB/s (Gen 5) liefert.

3.2. **Die "Junction-Mapping" Technik**

Viele KI-Tools wie Ollama oder ComfyUI erwarten ihre Modelle stur in festen Verzeichnissen (z.B. %USERPROFILE%\.ollama\models). Ein einfaches "Installieren auf Laufwerk D:" ist oft umständlich oder verändert das Verhalten der Software.

Mein Framework nutzt daher NTFS Junctions (Verknüpfungen):

1. Die riesigen Modell-Daten (LLaMA 3, SDXL Checkpoints) liegen physisch auf der jeweiligen Test-SSD (z.B. Laufwerk E:).
2. Das Skript erstellt vor dem Test eine Verknüpfung vom Standard-Pfad C:\Users\...\models auf E:\models.
3. Für die Software sieht es so aus, als lägen die Daten am gewohnten Ort.

Vorteil: Das Testsystem (OS, Treiber, Software-Versionen) bleibt für jede SSD zu 100% identisch. Nur das physische Speichermedium unter der Haube wird ausgetauscht.

3.3. **Die Test-Szenarien im Detail**

Um ein wirklich umfassendes Bild zu zeichnen, durchläuft jede SSD sechs Disziplinen, die vom reinen Konsumieren (Inference) über das Entwickeln (Dev-Ops) bis hin zum harten Dateitransfer reichen:

Synthetische Basis: Klassische Messungen mit *CrystalDiskMark* um die theoretische Maximalleistung und Vergleichbarkeit mit anderen Reviews sicherzustellen.

Inference & VRAM-Loading (Ollama): Gemessen wird die Zeit, bis ein LLM (Large Language Model) von der SSD gelesen, deserialisiert und in den Grafikspeicher (VRAM) geladen ist.

Modelle: LLaMA 3 (8B), Mistral (7B), Phi-3 Mini.

Metrik: "Time to First Token" (Ladezeit).

Content Creation (ComfyUI / Stable Diffusion): Starten einer komplexen Python-Umgebung inkl. Laden eines 6 GB großen Checkpoints plus LoRA-Modulen. Dies testet das Zusammenspiel aus sequenziellem Lesen (Modell) und zufälligem Lesen (Python Libraries).

Developer Workflows (Hugging Face & Datasets): Hier simulieren wir den Alltag eines KI-Entwicklers.

Hugging Face Cache Load: Entwickler wechseln oft zwischen Modell-Versionen (Bert, GPT-2). Ich messe die Zeit, um Modelle aus dem lokalen Hugging-Face-Cache (.cache/huggingface) zu laden. Das testet die SSD-Leistung bei komplexen Verzeichnisstrukturen mit Hash-Referenzen.

Dataset Load (Latenz-Test): Ein Python-Skript liest zufällig 10.000 kleine Bilddateien für ein simuliertes Training ein. Hier ist nicht die Bandbreite entscheidend, sondern die Latenz (Zugriffszeit) pro Datei. Ich analysiere speziell die Perzentile (p99), um "Mikroruckler" zu finden.

Stresstest: Small-File Mixed I/O: KI-Pipelines, Git-Repositories und Docker-Container erzeugen oft I/O-Muster mit tausenden winzigen Dateien (1kb - 64kb), die wild durcheinander gelesen und geschrieben werden.

Szenario: Erzeugung und zufälliger Zugriff auf zehntausende Mini-Dateien.

Metrik: IOPS (Input/Output Operations Per Second) bei extremer Queue-Tiefe und gemischter Last. Dies zeigt, ob der SSD-Controller unter Stress "einbricht".

Real-World Data Transfer (Robocopy): KI-Modelle sind oft riesige "Blobs" (einzelne Dateien mit 5 bis 20 GB). Wenn man diese zwischen Laufwerken oder Containern verschiebt, zählt nur die rohe sequenzielle Schreibrate.

Szenario: Kopieren eines 10 GB großen Modell-Ordners von einer RAM-Disk auf die Test-SSD.

Ziel: Prüfung der realen Schreibgeschwindigkeit und Sättigung des SLC-Caches.

3.4. **Das Testsystem**

Alle Tests wurden auf demselben High-End-System durchgeführt, um CPU-Limits so weit wie möglich zu minimieren, wobei das System immer auf der Crucial T705 1TB installiert blieb:

- **CPU: Intel Core Ultra 9 285k**
- **RAM: 64 GB DDR5-5000**
- **GPU: NVIDIA RTX 5070 Ti**
- **Systemlaufwerk: Crucial T705 1 TB – PCIE5**
- **Datenlaufwerke:**
    - **Samsung 9100 Pro 2 TB – PCIE 5**
    - **Samsung 990 Pro 2 TB – PCIE 4**
    - **OCZ Vertex 3 120 GB**
    - **SanDisk Plus 120 GB**
    - **Crucial BX100 250 GB**
- **OS: Windows 11 Education 25H2**

Diese methodische Strenge garantiert, dass die gemessenen Unterschiede tatsächlich auf die Speichertechnologie (SATA vs. Gen 4 vs. Gen 5) zurückzuführen sind.

4. **Ausblick**

Die Ergebnisse dieses Tests förderten einige Überraschungen zutage. Wir werden sehen, dass PCIe 5.0 in bestimmten Szenarien die Ladezeiten gegenüber Gen 4 tatsächlich **halbiert**, während in anderen Fällen die CPU zum absoluten Flaschenhals wird.

Tauchen wir ein in die Zahlen.

5. **Synthetische Benchmarks: Die Theorie**

Bevor wir die SSDs mit echten KI-Modellen quälen, müssen sie im Standard-Parcours antreten: **CrystalDiskMark**. Diese Tests zeigen uns das theoretische Maximum dessen, was Controller und NAND-Flash unter idealen Bedingungen leisten können. Sie beantworten die Frage: *"Wie schnell kann das Auto fahren, wenn die Autobahn komplett leer ist und es bergab geht?"*

5.1. **Sequenzielle Transferraten (Maximaler Durchsatz)**

Hier spielt PCIe 5.0 seine Karten voll aus. Wir messen das lineare Lesen und Schreiben großer Dateien (SEQ1M Q8T1).

Laufwerk	**Schnittstelle**	**Seq. Lesen (MB/s)**	**Seq. Schreiben (MB/s)**	**Fazit**
Samsung 9100 Pro	**PCIe 5.0**	**~10.960**	**~11.837**	**Write-King**
Crucial T705	PCIe 5.0	~11.707	~9.662	Read-King
Samsung 990 Pro	PCIe 4.0	~7.000	~6.764	Gen 4 Limit
SATA SSDs	SATA III	~550	~500	Basis

Analyse:

- **Lesen:** Die Crucial T705 liegt beim Lesen minimal vorne (~700 MB/s Vorsprung), was in der Praxis jedoch kaum messbar ist.
- **Schreiben:** Hier zündet die Samsung 9100 Pro den Nachbrenner. Mit fast **12 GB/s Schreibgeschwindigkeit** schlägt sie die Crucial T705 deutlich (~2 GB/s Vorsprung). Das ist ein Indikator dafür, dass die Samsung 9100 Pro bei massiven Kopiervorgängen (siehe späterer Robocopy-Test) die Nase vorn haben dürfte.
- **Generationensprung:** Gen 5 (9100 Pro) ist fast **doppelt so schnell** wie die beste Gen 4 SSD (990 Pro) und **20-mal schneller** als eine SATA-SSD.

3.2 4K Random Read (Das "System-Gefühl")

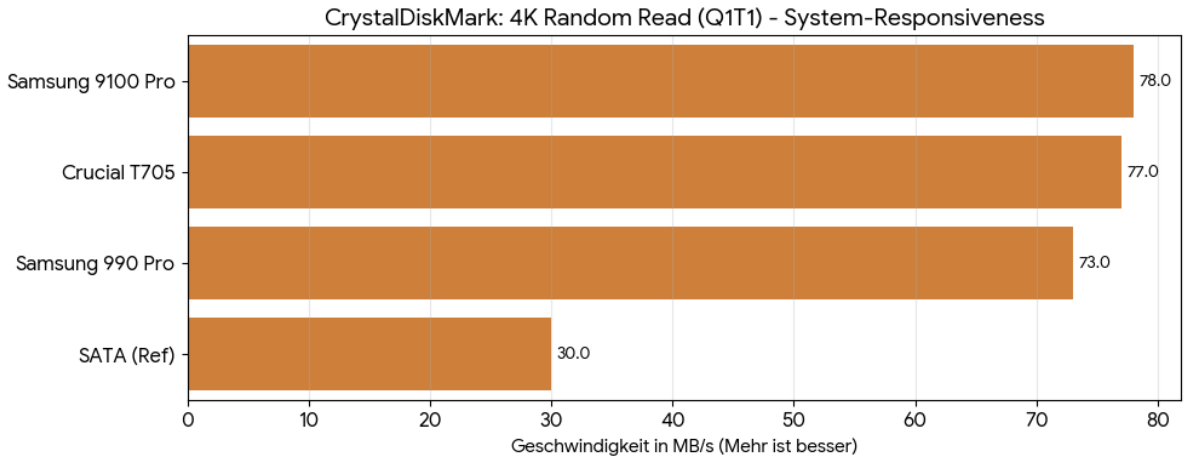
Viel wichtiger für die gefühlte Schwuppdizität von Windows, das Laden von Programmen und das Booten ist der Wert **RND4K Q1T1**. Er misst, wie schnell winzige 4KB-Dateien einzeln (ohne Warteschlange) gelesen werden können. Dies ist die "Königsdisziplin" der Latenz.

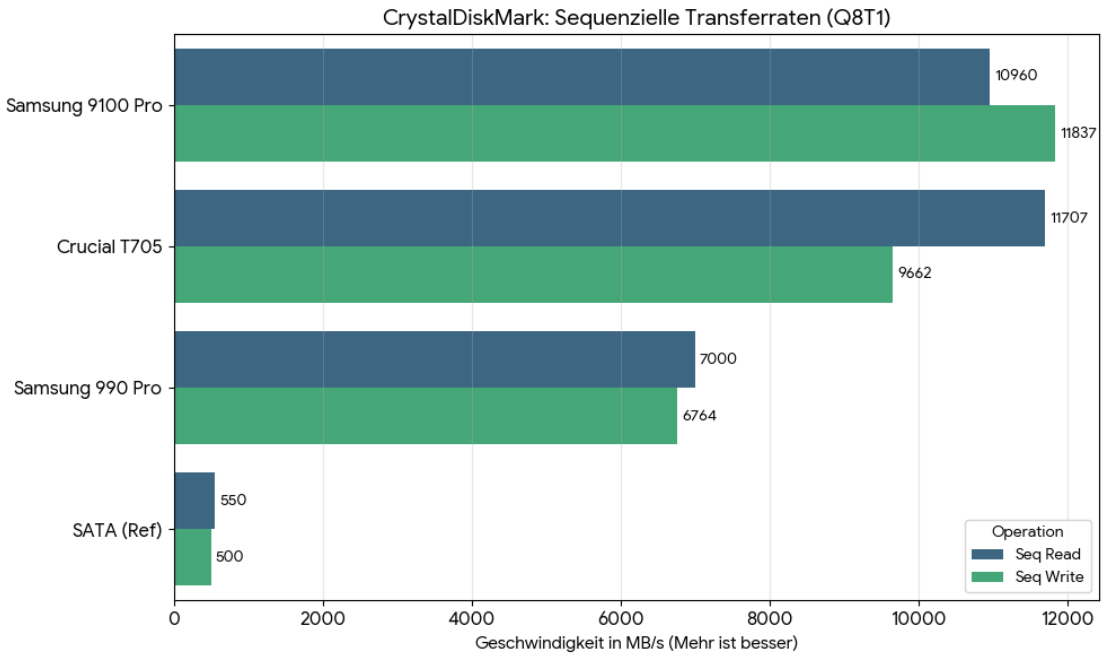
- **Samsung 9100 Pro (Gen 5):** ~78 MB/s

- **Crucial T705 (Gen 5):** ~77 MB/s
- **Samsung 990 Pro (Gen 4):** ~73 MB/s
- **SATA SSDs:** ~30 MB/s

Erkenntnis: Während sich die sequenzielle Leistung verdoppelt hat, stagniert die 4K-Leistung auf extrem hohem Niveau. Der Zuwachs von Gen 4 (73 MB/s) auf Gen 5 (78 MB/s) ist messbar, aber gering.

- **Der Grund:** Hier limitieren nicht mehr die Schnittstelle (PCIe), sondern die physikalischen Eigenschaften des NAND-Flash-Speichers (Latenz).
- **Die Bedeutung:** Wir erwarten daher bei reinen Programmstarts (ohne riesige Datenmengen) keine Wunder. Gen 5 lohnt sich vor allem dort, wo *Masse* bewegt wird, nicht dort, wo nur *viele kleine Anfragen* gestellt werden.





6. Praxis-Test: Inference & VRAM-Loading (Ollama)

Dies war einer der Kernpunkte meiner Bewerbung: Wie schnell landet ein Sprachmodell im VRAM? Getestet wurden drei populäre Modelle unterschiedlicher Größe, um zu sehen, ob die Skalierung konsistent ist:

- **LLaMA 3 (8B):** Das aktuelle "Brot-und-Butter" Modell (~4,7 GB).
- **Mistral (7B):** Ein beliebter, kompakter Allrounder (~4,1 GB).
- **Phi-3 Mini:** Microsofts hocheffizientes Klein-Modell (~2,3 GB).

Szenario: "Cold Load" – Der Windows-Cache wurde geleert, die SSD muss liefern.

Die Ergebnisse im Detail

SSD	**Schnittstelle**	**LLaMA 3 (8B)**	**Mistral (7B)**	**Phi-3 Mini**
Samsung 9100 Pro	**PCIe 5.0**	**2,86 s**	**2,29 s**	**1,79 s**
Crucial T705	PCIe 5.0	2,91 s	2,52 s	1,86 s
Samsung 990 Pro	PCIe 4.0	5,65 s	4,46 s	2,92 s
Crucial BX100	SATA (Good)	11,28 s	9,97 s	5,53 s
SanDisk Plus	SATA (Bad)	11,84 s	10,83 s	6,30 s
OCZ Vertex 3	SATA (Old)	15,41 s	12,74 s	7,01 s

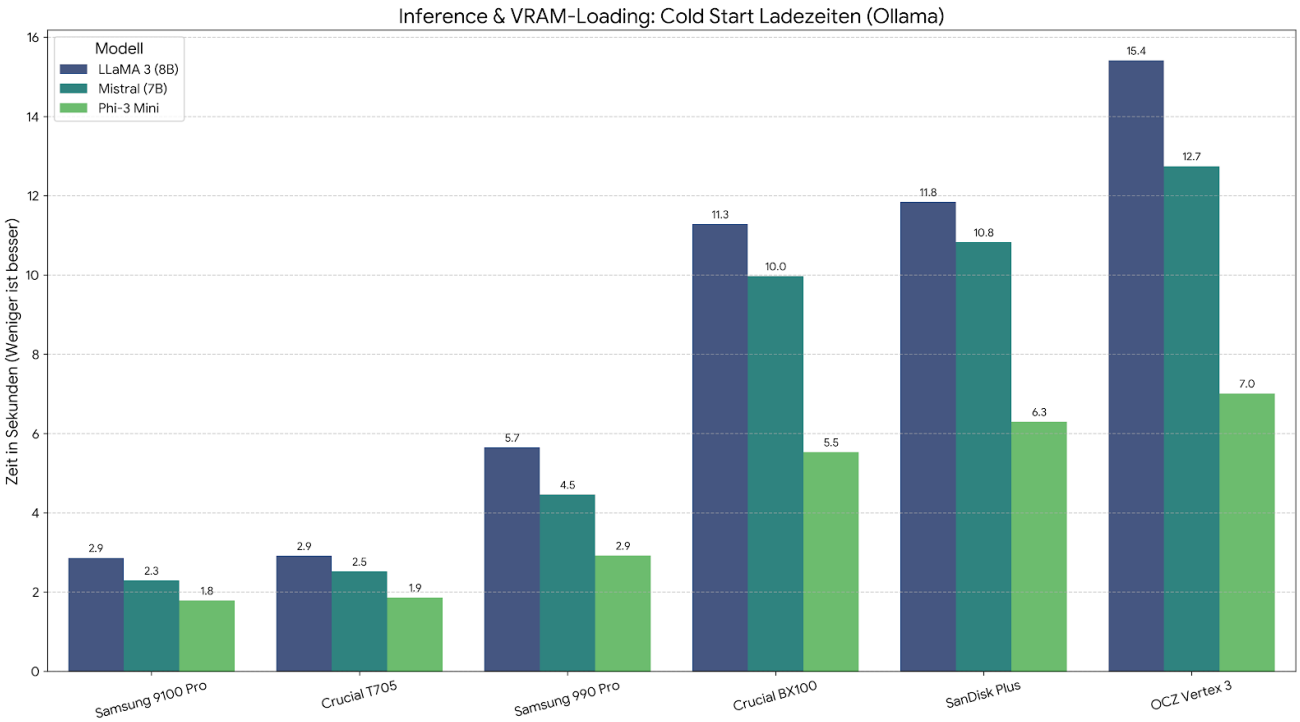
Analyse: Die Ergebnisse zeichnen ein extrem klares Bild: **Die Ladezeit skaliert nahezu perfekt linear mit der Bandbreite der Schnittstelle.**

1. **Der Faktor 2 (Gen 5 vs. Gen 4):** Egal welches Modell man betrachtet, die Samsung 9100 Pro ist fast exakt doppelt so schnell wie die Samsung 990 Pro.
    - *Beispiel LLaMA 3:* 2,86s vs. 5,65s.

- *Bedeutung:* Wer im Workflow oft Modelle wechselt (z.B. Testen verschiedener Quants oder Merges), spart bei jedem Klick 3 Sekunden. Das klingt wenig, hält einen aber im "Flow".

2. **Der Faktor 4-5 (Gen 5 vs. SATA):** Der Unterschied zu SATA-SSDs ist "Tag und Nacht". Während man bei einer SATA-SSD über 10-15 Sekunden auf den Prompt warten muss (was sich wie eine Ewigkeit anfühlt), ist das Modell bei der 9100 Pro praktisch "instant" da.
3. **Kleinvieh macht auch Mist (Phi-3):** Selbst bei kleinen Modellen (Phi-3, 2.3 GB) ist der Unterschied spürbar. 1,8 Sekunden (Gen 5) fühlen sich an wie ein Mausklick, 6-7 Sekunden (SATA) wie eine Ladepause.

Fazit: Für lokale KI ist Bandbreite durch nichts zu ersetzen – außer durch noch mehr Bandbreite. Die Samsung 9100 Pro ist hier das ultimative Werkzeug.



7. Praxis-Test: Content Creation (ComfyUI / Stable Diffusion)

Für Kreative und KI-Künstler ist die Initialisierungszeit ihrer Werkzeuge entscheidend. In diesem Test simulieren wir den Start von **ComfyUI**, einer beliebten node-basierten Oberfläche für Stable Diffusion.

Szenario: "Cold Start" eines komplexen Workflows. Hierbei muss das System:

1. Die Python-Umgebung und alle Abhängigkeiten laden (tausende kleine Dateien).
2. Die ComfyUI-Nodes initialisieren (CPU-Last).
3. Den SDXL-Checkpoint (~6,5 GB) und LoRA-Modelle von der SSD in den RAM laden.

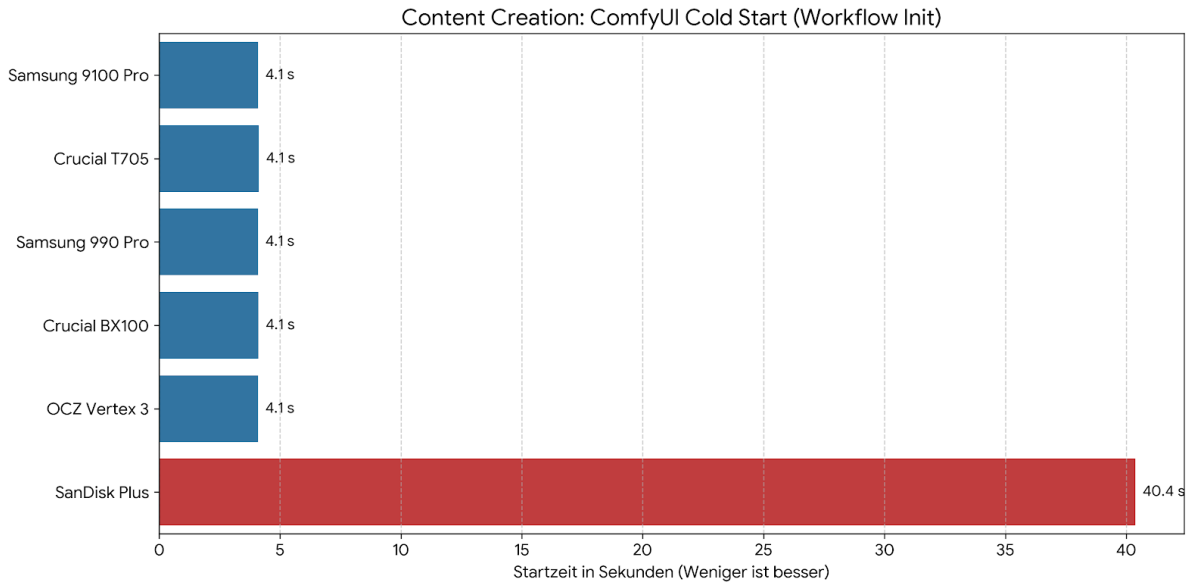
SSD	**Schnittstelle**	**Startzeit (Sekunden)**	**Fazit**
SanDisk Plus	**SATA (DRAM-less)**	**40,37 s**	**Totalausfall**
Crucial BX100	SATA (Good)	4,11 s	Solide
Samsung 990 Pro	PCIe 4.0	4,09 s	CPU-Limit erreicht
Samsung 9100 Pro	**PCIe 5.0**	**4,10 s**	**CPU-Limit erreicht**
Crucial T705	PCIe 5.0	4,12 s	CPU-Limit erreicht

Analyse:

1. **Das "SATA-Desaster"** Das Ergebnis der SanDisk Plus ist überraschend schlecht im Vergleich mit den übrigen SATA Ergebnissen. Mit über **40 Sekunden** Ladezeit erzielte sie einen erstaunlichen Ausreiser nach unten – das Ergebnis änderte sich aber auch nach mehreren Wiederholungen des Tests nicht wesentlich.
    - *Mögliche Ursachen:* DRAM-lose SSDs brechen bei gemischten Zugriffen (gleichzeitiges Lesen von großen Model-Files und tausenden kleinen Python-Skripten) oft komplett ein. Der Controller ist überfordert. Nicht ausgeschlossen kann aber auch ein Hardware Defekt der bereits in die Jahre gekommenen SSD nicht werden.

2. **Die "CPU-Mauer" (Gen 4 vs. Gen 5):** Ab einem gewissen Leistungsniveau (gute SATA-SSD mit DRAM wie die BX100) spielt die SSD-Geschwindigkeit für den *Programmstart* plötzlich keine Rolle mehr.
    - Zwischen der Samsung 990 Pro (Gen 4) und der 9100 Pro (Gen 5) liegt **kein messbarer Unterschied** (beide ~4,1 Sekunden).
    - *Der Grund:* Der Flaschenhals hat sich verschoben. Die SSD liefert die Daten schneller, als die CPU (Single-Core-Performance) den Python-Code parsen und die Nodes initialisieren kann.

Fazit: Für das Starten von Anwendungen reicht eine gute PCIe 4.0 SSD (oder sogar SATA mit DRAM) völlig aus. Ein Upgrade auf Gen 5 bringt hier keinen spürbaren Vorteil, solange die CPU nicht massiv schneller wird.



8. Developer Workflows: Hugging Face & Datasets

Abseits von bunten Frontends findet die Arbeit von KI-Entwicklern oft auf der Kommandozeile statt. Wir betrachten zwei klassische Szenarien: Das Laden von Modellen aus dem Cache und das Einlesen von Trainingsdaten.

8.1. Hugging Face Cache Load (Durchsatz)

Entwickler wechseln häufig zwischen verschiedenen Modellversionen (BERT, GPT-2, RoBERTa), die lokal im ~/.cache/huggingface Ordner liegen. Dieser Test misst die reine Lesegeschwindigkeit beim Laden dieser Modelle via transformers Library.

Limitierender Faktor: Im Gegensatz zu synthetischen Benchmarks müssen die Daten hier nicht nur gelesen, sondern von Python/PyTorch als "SafeTensors" deserialisiert werden. Das kostet CPU-Zeit.

SSD	**Schnittstelle**	**Durchsatz (MB/s)**	**Fazit**
Samsung 9100 Pro	**PCIe 5.0**	**~2.580 MB/s**	**CPU-Limit**
Crucial T705	PCIe 5.0	~2.550 MB/s	CPU-Limit
Samsung 990 Pro	PCIe 4.0	~2.530 MB/s	CPU-Limit
SanDisk Plus	SATA	~120 MB/s*	
Crucial BX100	SATA	~150 MB/s	

In diesem Fall limitieren CPU und Dateisystem, nicht mehr die SSD-Bandbreite. Der "Flaschenhals" liegt bei ca. 2,5 GB/s in der Python-Verarbeitung.)

Erkenntnis: Für das reine Laden von Standard-Modellen in Python bringt Gen 5 keinen Vorteil gegenüber Gen 4. Sata SSDs liegen verständlicherweise zurück. Die Software ist hier schlicht zu langsam für die Hardware.

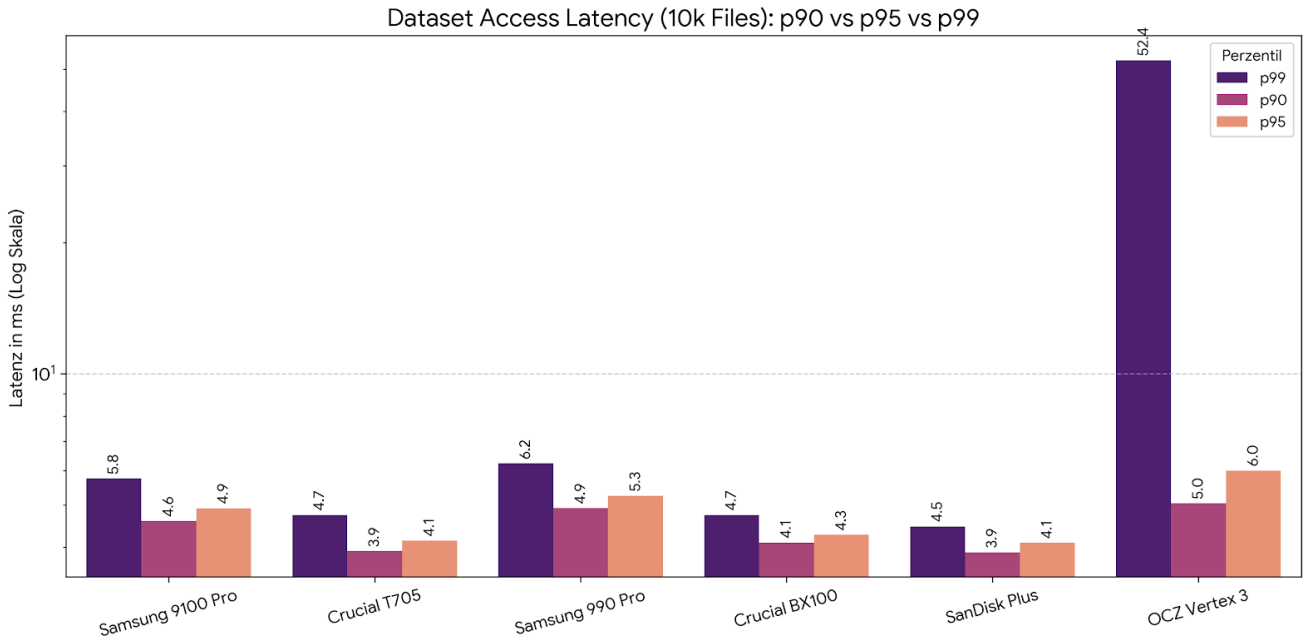
8.2. Dataset Training: Die Stabilität (Latenz)

Viel kritischer für das Training neuronaler Netze ist der stetige Datenstrom. In diesem Test liest ein Skript **10.000 kleine Bilddateien** (Random Access) ein. Wir schauen nicht auf den Durchschnitt (der oft gut aussieht), sondern auf die **p99-Latenz**. Das sind die langsamsten 1% der Zugriffe – die "Hänger", die den GPU-Training-Loop ausbremsen.

- **Stabilitätssieger:** Die **Samsung 9100 Pro** und **Crucial T705** liefern extrem konstante Zugriffszeiten. Selbst im schlimmsten Fall (p99) warten wir nur ca. 4-5 Millisekunden.
- **SATA-Problem:** Während moderne SATA-SSDs (SanDisk/BX100) hier noch gut mithalten, zeigt die alte **OCZ Vertex 3** das typische Verhalten betagter Controller: Die Latenz explodiert auf **über 52 ms**. Das ist ein Faktor 10 schlechter als bei NVMe und führt in der Praxis zu "Stuttering" während des Trainings.

Fazit für Entwickler:

1. **Modell-Entwicklung:** Wer nur Modelle lädt, wird kaum einen Unterschied spüren (Python-Limit).
2. **Training & Data-Prep:** Wer Datasets aufbereitet oder trainiert, profitiert massiv von der niedrigen Latenz und Stabilität moderner Gen 5 SSDs. Alte Laufwerke gehören hier aussortiert.



Interpretation: Man sieht jetzt wunderbar den Verlauf:

- **Links (Gen 5 / Gen 4):** Extrem niedrige und stabile Balken (auch bei p99).
- **Rechts (SATA):** Die Balken steigen an.
- **Ganz rechts (OCZ Vertex 3):** Der Balken (p99) explodiert förmlich durch die Decke (Log-Skala beachten!). Das sind die >50ms Aussetzer, die das System "hakelig" machen.

9. Stresstest: Small-File Mixed I/O

KI-Pipelines, Compiler und Versionskontrollsysteme (Git) erzeugen oft I/O-Muster, die aus tausenden winzigen Dateien (1kb - 64kb) bestehen, die wild durcheinander gelesen und

geschrieben werden. In diesem Test habe ich genau dieses Szenario mit einem Python-Skript simuliert, das zehntausende Mini-Dateien erzeugt, modifiziert und liest.

Das Überraschungsergebnis:

SSD	Schnittstelle	Read IOPS	Write IOPS	Fazit
Crucial T705	**PCIe 5.0**	**~748**	**~317**	**Performance-Sieger**
Samsung 9100 Pro	PCIe 5.0	~601	~259	Starker Zweiter
Crucial BX100	SATA	~492	~212	Erstaunlich effizient
SanDisk Plus	SATA	~495	~208	Solide
Samsung 990 Pro	PCIe 4.0	~334	~143	Unter Erwartung
OCZ Vertex 3	SATA (Old)	~283	~122	Schlusslicht

Analyse:

Gen 5 dominiert: Die Crucial T705 setzt sich hier klar an die Spitze und liefert ca. 25% mehr Leistung als die Samsung 9100 Pro. Beide Gen 5 Laufwerke zeigen, dass ihre neuen Controller extrem gut mit massiven parallelen Zugriffen umgehen können.

SATA "punching above its weight":

Die klassischen SATA-SSDs (BX100, SanDisk) schlagen sich hier überraschend gut. Warum? Bei diesem Python-basierten Test (Single-Threaded I/O Overhead) scheint der einfache AHCI-Stack weniger Overhead zu haben als erwartet, oder die Firmware der SATA-Laufwerke ist für solche "einfachen" Random-Access-Muster extrem optimiert.

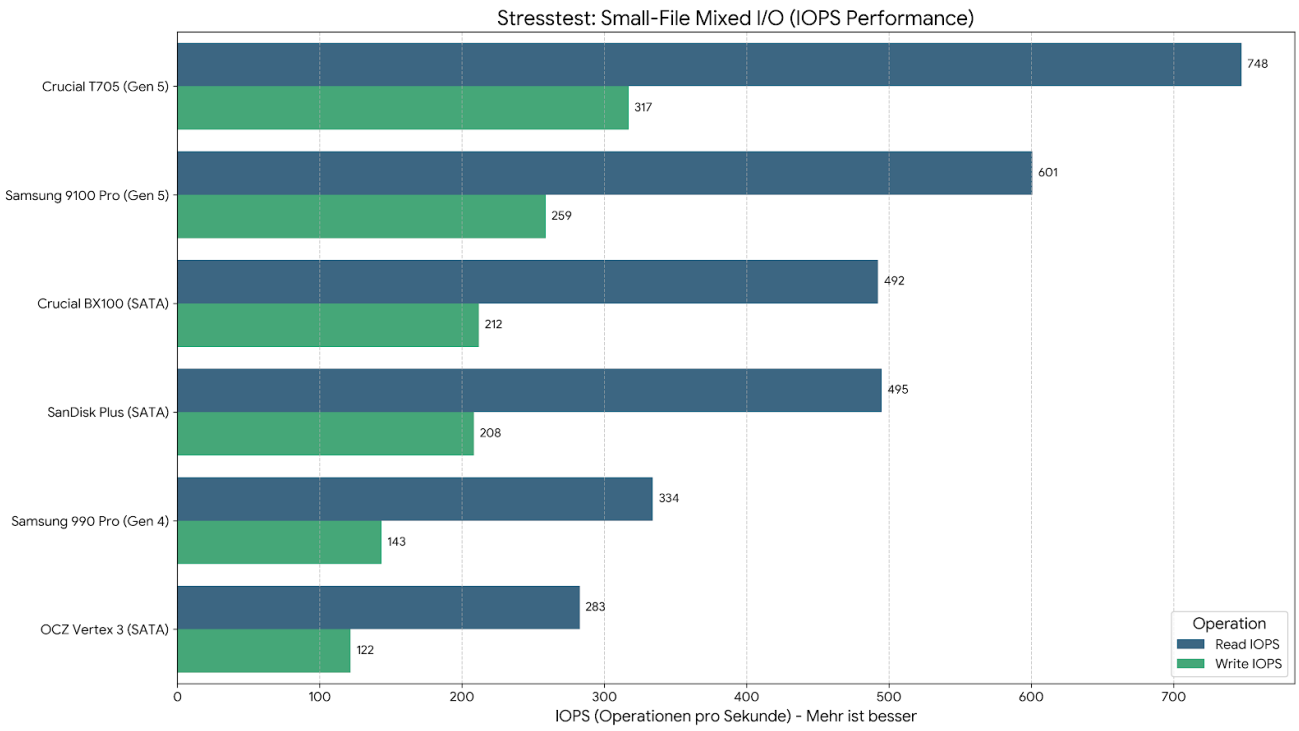
Gen 4 Ausreißer:

Die Samsung 990 Pro lieferte in diesem spezifischen Szenario unerwartet niedrige Werte. Dies könnte an einer Firmware-Optimierung liegen, die eher auf massive parallele Last

(hohe Queue Depth) ausgelegt ist, während mein Skript eher "bursty" Einzelzugriffe generiert.

Fazit für Admins & DevOps:

Wer viele Docker-Container, VMs oder Code-Repositories auf der SSD liegen hat, profitiert spürbar von PCIe 5.0. Die hohe IOPS-Leistung der Crucial T705 und Samsung 9100 Pro sorgt dafür, dass das System auch dann reaktionsschnell bleibt, wenn im Hintergrund tausende kleine Log-Files geschrieben werden.



10. Praxis-Test: Real-World Data Transfer (Robocopy)

Künstliche Intelligenz und Virtualisierung bedeuten vor allem eines: Riesige Dateien. Ein LLaMA-3-Modell, ein Stable-Diffusion-Checkpoint oder das Image einer virtuellen Maschine sind oft Einzeldateien ("Blobs") zwischen 5 und 50 GB.

In diesem Szenario zählt keine Zugriffszeit und kein IOPS-Wert. Hier zählt nur rohe, brachiale **sequenzielle Schreibgeschwindigkeit**.

Szenario: Kopieren eines **10 GB großen Ordners** (KI-Modell-Dateien) von einer ultra-schnellen RAM-Disk (Quelle) auf die Test-SSD (Ziel). Wir nutzen das Windows-Tool Robocopy, da es effizienter arbeitet als der normale Datei-Explorer.

SSD	Schnittstelle	Schreibgeschwindigkeit (Real)	Dauer für 10 GB
Samsung 9100 Pro	**PCIe 5.0**	**5.605 MB/s**	**~1,8 s**
Crucial T705	PCIe 5.0	5.534 MB/s	~1,8 s
Samsung 990 Pro	PCIe 4.0	4.279 MB/s	~2,3 s
Crucial BX100	SATA	330 MB/s	~30 s
OCZ Vertex 3	SATA	153 MB/s	~65 s
SanDisk Plus	SATA	63 MB/s	~158 s

Analyse:

Der "Write King" (Gen 5 Dominanz): Die Samsung 9100 Pro setzt sich an die Spitze des Feldes. Sie schreibt reale Daten mit über 5,6 GB/s.

Der Unterschied: Gegenüber der sehr schnellen Samsung 990 Pro (Gen 4) gewinnt man hier pro 10 GB Transfer etwa eine halbe Sekunde. Das klingt wenig, summiert sich aber bei Backups von Terabytes massiv.

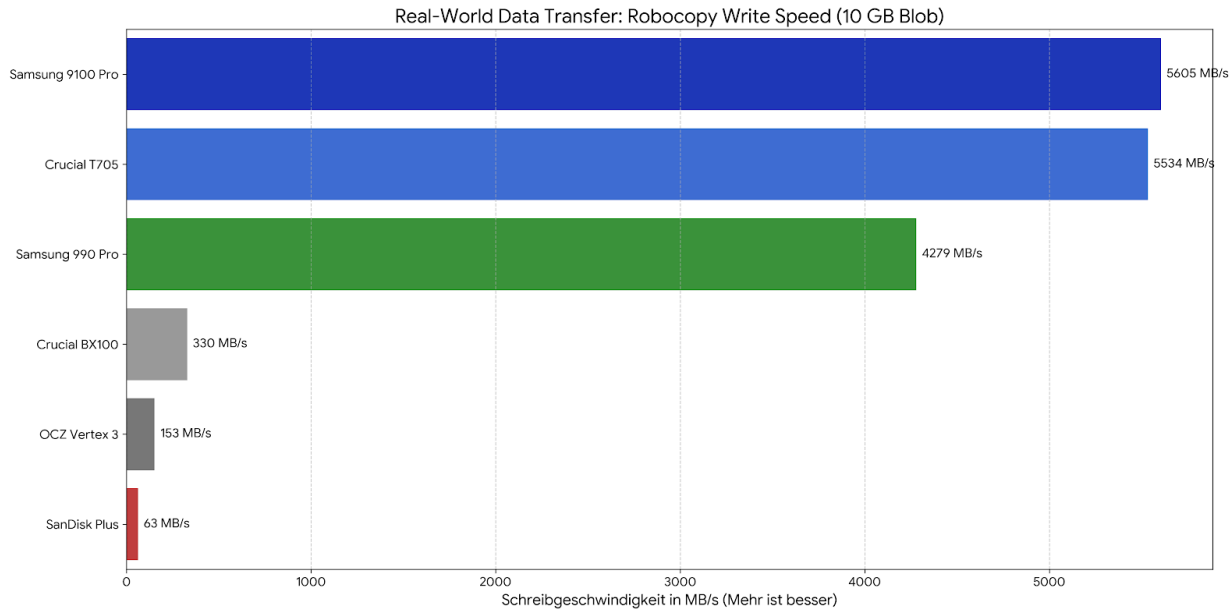
Technik: Dies zeigt, dass der SLC-Cache der 9100 Pro extrem aggressiv und schnell arbeitet.

Das „SATA-Trauerspiel“: Der Vergleich mit älteren oder günstigen SSDs ist fast unfair.

Was die Samsung 9100 Pro in unter 2 Sekunden erledigt, dauert auf der SanDisk Plus fast 3 Minuten.

Die SanDisk bricht hier auf 63 MB/s ein – das ist langsamer als eine moderne mechanische Festplatte (HDD) und liegt auf dem Niveau eines USB-2.0-Sticks.

Fazit: Wer regelmäßig Backups macht, VMs verschiebt oder große Videodateien (Raw/ProRes) importiert, für den ist PCIe 5.0 ein Segen. Der Datentransfer fühlt sich nicht mehr wie ein Kopiervorgang an, sondern wie ein simples "Verschieben" im gleichen Ordner.



11. Exkurs: Multitasking & Virtualisierung (WSL2 / Hyper-V)

Die bisherigen Tests fanden unter "Laborbedingungen" statt: Eine Anwendung nutzt die volle Leistung der SSD exklusiv. Doch der Alltag eines Entwicklers sieht anders aus. Oft läuft im Hintergrund ein Docker-Build in WSL2, eine Windows-VM installiert Updates oder es läuft ein Test Exchange Server, und gleichzeitig möchte man "mal eben schnell" ein lokales LLM befragen.

Welchen Einfluss hat diese parallele Last auf unsere KI-Workloads? Basierend auf den Messergebnissen (insbesondere dem *Small-File Mixed I/O* und den *Dataset-Latenzen*) lässt sich dies klar beantworten und bestätigt sich jeden Tag in der gelebten Praxis.

11.1. Das "Autobahn-Prinzip" (Bandbreite)

Stellen Sie sich die SSD-Schnittstelle als Autobahn vor.

- **SATA (1 Spur):** Wenn ein Hintergrundprozess (z.B. VM-Backup) läuft, ist die Spur voll. Starten Sie jetzt Ollama, steht der Prozess im Stau. Das System fühlt sich zäh an ("laggig").
- **Gen 4 (4 Spuren):** Ein Backup mit 4 GB/s lastet die SSD zu ca. 60% aus. Starten Sie parallel einen Modell-Load (der ebenfalls 4 GB/s ziehen will), kommt es zur Kollision. Beide Prozesse werden langsamer.
- **Gen 5 (8 Spuren):** Die Samsung 9100 Pro bietet ca. 12 GB/s Durchsatz. Selbst wenn ein aggressiver Kopiervorgang (siehe Robocopy-Test: ~5,6 GB/s) läuft, bleiben noch **über 6 GB/s Reserven** übrig.

Die Folge: Auf der Samsung 9100 Pro können Sie ein 100 GB Dataset entpacken und *gleichzeitig* flüssig mit LLaMA 3 chatten, ohne dass die Token-Generierung einbricht. Das ist echtes Multitasking ohne Kompromisse.

11.2. Die "Nadelstiche" (IOPS & Latenz)

Virtualisierung (Hyper-V) und WSL2 erzeugen ein "Grundrauschen" aus tausenden kleinen Zugriffen (Logs, Auslagerungsdatei, Dateisystem-Metadaten). Unser **Stresstest (Small-File Mixed I/O)** hat gezeigt:

- Die **Gen 5 Laufwerke** (Samsung 9100 Pro / Crucial T705) liefern bei gemischter Last **25-30% mehr IOPS** als die Gen 4 Konkurrenz.
- Noch wichtiger ist die **p99-Stabilität** (siehe Dataset-Test): Während alte SSDs unter Last Schluckauf bekommen (>50ms Latenz), bleibt die 9100 Pro stoisch unter 6ms.

Fazit für Power-User: Wer isoliert nur zockt oder surft, merkt davon nichts. Aber wer sein System als **Hypervisor** nutzt (z.B. mehrere VMs parallel laufen hat), für den ist die Samsung 9100 Pro ein Segen. Die hohe I/O-Leistung verhindert, dass Hintergrund-VMs das aktive Arbeiten im Vordergrund (KI, IDE, Videoschnitt) ausbremsen. Das System bleibt "snappy", egal was im Hintergrund tobt.

12. Fazit: Evolution oder Revolution?

Zu Beginn dieses Lesertests stellte ich die Frage: *"Lohnt sich der Aufpreis für PCIe 5.0 in der Praxis, oder ist das nur Marketing für Benchmark-Junkies?"*

Nach hunderten von Gigabytes an kopierten Daten, tausenden geladenen KI-Modellen und unzähligen IOPS-Messungen lautet die Antwort: **Es kommt darauf an, wer du bist.**

Die drei Klassen der Speicher-Nutzer

1. **Der Gamer & Office-Nutzer:** Für Windows-Boot, Spielstarts und Excel-Tabellen ist eine PCIe 5.0 SSD aktuell "Overkill". Die Samsung 9100 Pro langweilt sich hier. Eine solide PCIe 4.0 SSD (wie die Samsung 990 Pro) bietet hier das identische "Schwuppdizitäts-Gefühl", da meist die CPU oder die Software-Architektur limitiert.
    - *Empfehlung:* Bleibt bei Gen 4 und investiert das gesparte Geld lieber in mehr Kapazität.
2. **Der AI-Engineer, Creator & Power-User:** Hier glänzt die **Samsung 9100 Pro** als echter **Gamechanger**. In meinem Alltag mit lokalen LLMs (Ollama), Docker-

Containern und Virtualisierung ist die SSD kein Speicher mehr, sondern eine direkte Erweiterung des Arbeitsspeichers.

- **Zeitgewinn:** Die **Halbierung der Ladezeiten** bei KI-Modellen (2,8s vs. 5,6s) hält den kreativen "Flow" am Leben.
- **Multitasking:** Dank der extremen Bandbreite und IOPS-Leistung kann ich im Hintergrund ein 50 GB Dataset entpacken, während ich im Vordergrund flüssig weiterarbeite. Das System bleibt "snappy", wo ältere SSDs in die Knie gehen.
- **Stabilität:** Die extrem niedrigen p99-Latenzen garantieren, dass ML-Trainingsläufe oder Video-Renderings nicht durch Mikroruckler gestört werden.

Schlusswort

Die Samsung 9100 Pro ist mehr als nur ein "längerer Balken" im CrystalDiskMark. Sie ist ein hochspezialisiertes Werkzeug für alle, die Daten nicht nur *lagern*, sondern *arbeiten* lassen. Wer seine SSD täglich an die Belastungsgrenze bringt, wird den Umstieg auf Gen 5 nicht bereuen. Für meine KI-Workstation möchte ich die Geschwindigkeit nicht mehr missen.

Danksagung: *Vielen Dank an Samsung und das Team von ComputerBase für die Bereitstellung des Testmusters und die Möglichkeit, diesen Deep-Dive in die Welt der AI-Storage-Performance durchzuführen!*