deepseek and the law

IP & Data Risks in Open-Source LLMs

What is 🐋 deepseek ?

Who & What:
- Founded 2023, Hangzhou (China), by hedge-fund entrepreneur Liang Wenfeng
- Open-source-weights LLM family: DeepSeek-V3 (general) & DeepSeek-R1 (reasoning)

Business Model:
- Paid, ultra-cheap APIs (~ $2.19 / M tokens versus OpenAI's up to $60),
- Custom enterprise builds
- Commercial integrations (Baidu, BMW, ...)

Why It Took Off:
- Free consumer chatbot drives brand awareness becoming top-ranked free chatbot app
- Hedge-fund: lets it undercut competitors on price and iterate fast
- Cheap training: Trained R1 for ≈ $6 M (compute costs), lowest among competitors (current token prices)
- Open source weights under an MIT-based licence to rapidly build a strong developer community ·

SCRAPING ISSUES

OPEN WEIGHTS BUT SCRAPED TRAINING DATA SECRET

InfoSoc Dir 2001/29

- **Art 2:** bulk scraping = permanent reproduction.

- **Art 3:** outputs with substantial original work = "communication to the public."

- **17 Feb 2025:** ICMP accused DeepSeek of full-lyrics scraping.

- **Art 8 (3)** allows EU-wide injunctions against intermediaries.

Database Dir 96/9

- **Art 3:** copying a database structure infringes copyright.

- **Art 7:** sui-generis right bars extraction of a "substantial part";

- **C-30/14 Ryanair v PR Aviation** confirms that website terms & conditions can validly forbid scraping, even where the database itself lacks EU protection.

CDSM Dir 2019/790

- **Art 3:** applies only to research organisations, so DeepSeek cannot rely on it

- **Art 4:** for-profit TDM only if data was lawfully accessible and the owner did not opt-out (robots.txt/metadata).

- One ignored opt-out can trigger an injunction as in NY Times vs OpenAI.

SCRAPING ISSUES
OPEN WEIGHTS BUT SCRAPED TRAINING DATA SECRET

AI Act 2024/1689

- **Art 10:** training data must be lawful, relevant, accurate—illegal or biased scraping breaches this duty.

- **Arts 52-53:** providers must publish data-source list, copyright-compliance policy, risk mitigations.

- **Arts 22-54:** non-EU providers need an EU representative before release.

GDPR – Regulation (EU) 2016/679

Scraped personal data raises separate consent, purpose-limitation and erasure obligations—analysed further in the dedicated section.

DISTILLATION CASE:
OPENAI VS DEEPSEEK

Background
OpenAI's Statement to the New York Times

*"We are aware of and reviewing indications that DeepSeek may have inappropriately **DISTILLED** our models, and will share information as we know more. We take aggressive, proactive countermeasures to protect our technology and will continue working closely with the US government to protect the most capable models being built here."*
(Source: NYT, 2025)

Transparency issues
OpenAI - Deepseek comparison
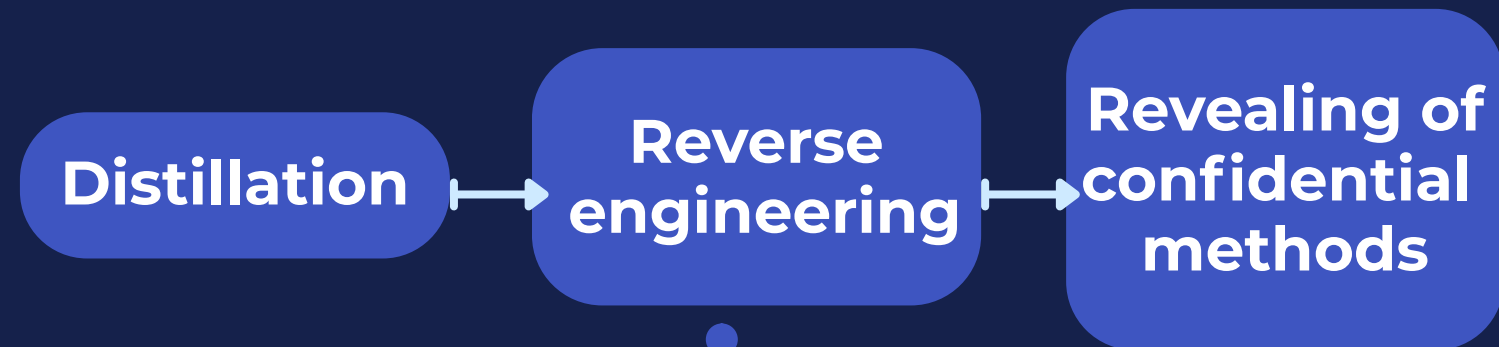No information about data sources from DeepSeek

Evidences
Copyleaks
74.2% of DeepSeek-R1's outputs closely resembled those of OpenAI's ChatGPT

LEGAL ISSUES

Trade secret misappropriation

Distillation → Reverse engineering → Revealing of confidential methods

✅ **Lawful:**

"**independent** discovery or creation" Art 3.1(a) TSD

❌ **Unlawful:**

"*being in breach of a **contractual** or any other **duty** to **limit** the **use of the trade secret**.*"
Art 4.3(c) TSD

"*observation, study, **disassembly** or testing of a product … that has been made **available to the public** … **free** from any legally valid **duty** to **limit the acquisition of the trade secret***"
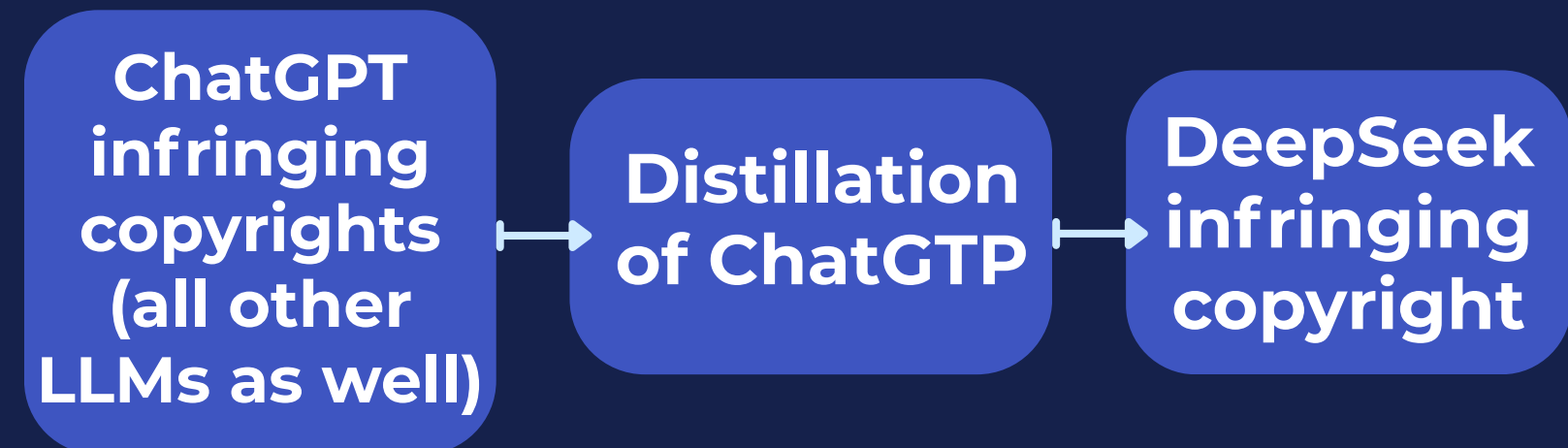Art 3.1(b)

Copyright infringement

Are LLMs violating copyright ?

Answer: OpenAI v. New York Times

But for DeepSeek there is more…
Distillation-based copyright infringement

ChatGPT infringing copyrights (all other LLMs as well) → Distillation of ChatGTP → DeepSeek infringing copyright

Terms of service prohibit to:

(d) reverse assemble, reverse compile, decompile, translate, engage in **model extraction** or **stealing attacks**, or otherwise attempt to **discover the source code** or **underlying components** of the Services, **algorithms**, and **systems** of the Services

(e) *use Output to **develop any artificial intelligence models** that **compete** with our products and services.*

TRADEMARKS AND BRANDING RISKS

TRADEMARK RISKS: DEEPSEEK VS OPENAI

▶ **Risk of and Confusion**
Users might mistakenly believe DeepSeek is affiliated with OpenAI.

▶ **Look-and-Feel Imitation**
OpenAI could argue that DeepSeek mimics ChatGPT's style, structure, and behaviour, misleading users.

▶ **Strategic Risks**
- Market trust could shift based on perceived affiliation
- Potential for legal claims or platform bans (app stores, APIs)

▶ **Defensive Strategies for DeepSeek**
Rebranding, clear disclaimers

PATENTS AND
FREEDOM-TO-OPERATE (FTO)
PATENT RISKS: NO FTO ANALYSIS

▶ **Weak Patent Position Outside China, no international patents**
DeepSeek owns patents only in China → No international protection in EU or US.

▶ **No Freedom-To-Operate (FTO) analysis**
Without FTO due diligence, DeepSeek risks infringing third-party patents on: RLHF training methods, Model Routing algorithms and Quantization techniques

▶ **EU patent law (EPC Art. 52–56):** need for technical effect + inventiveness
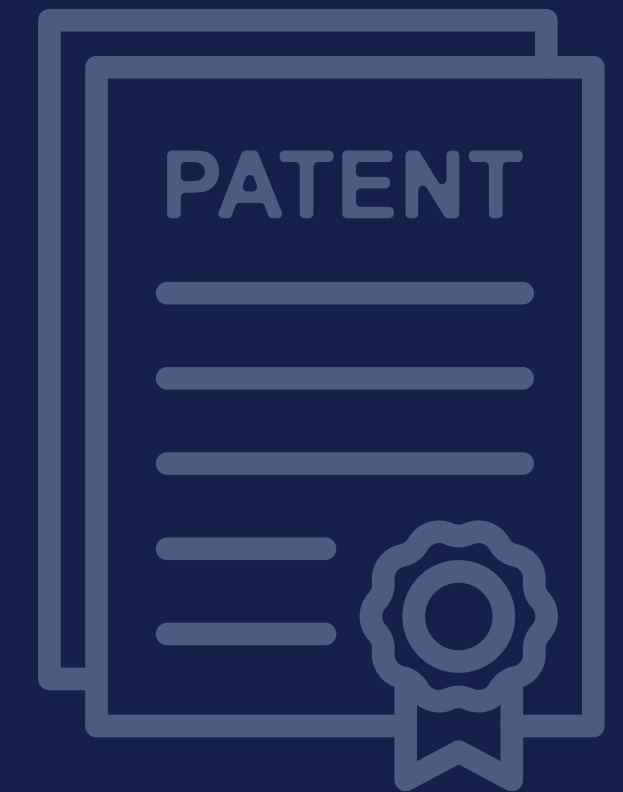
What DeepSeek Should Do
- Conduct urgent FTO analysis before EU market expansion
- Design-around patented methods where necessary
- Consider licensing negotiations if infringement is unavoidable

PATENT

GDPR VIOLATIONS

DEEPSEEK UNDER SCRUTINY

24 Jan 2025 → 29 Jan 2025 Garante requests full GDPR dossier; DeepSeek denies GDPR applicability (Art. 3), fails to cooperate (Art. 31), and lacks adequate security measures (Art. 32).
30 Jan 2025 Regulator issues an emergency order blocking the app and limiting all processing in Italy because of "totally insufficient" reply

No legal basis for processing **Art. 6 GDPR:**
DeepSeek collected data without valid consent or justification

No age verification for minors **Art. 8 GDPR**
The app lacks mechanisms to prevent access by underage users

Lack of transparency **Art. 12–14 GDPR**
No clear privacy notices on how personal data is used or stored

No EU Representative **Art. 27 GDPR**
Mandatory under GDPR for non-EU providers — DeepSeek failed to appoint one

The Italian DPA suspended DeepSeek's data processing in Italy (Art. 58). If unresolved, this may lead to a permanent ban and significant fines (Art. 83).
The case prompted investigations in **France, Belgium, and Ireland**.
As of April 2025, DeepSeek remains limited and under EU regulatory scrutiny.

EXPORTING DATA IMPORTING RISK

VIOLATING GDPR BEYOND EU BORDERS

- DeepSeek processes user data and stores logs on servers located in China — a country without an EU adequacy decision
- No Standard Contractual Clauses, Binding Corporate Rules, or other valid legal mechanisms were implemented, in breach of **Articles 44–49** of the GDPR
- Under the Schrems II ruling, such transfers require **rigorous risk assessments,** particularly when directed to jurisdictions like China, where expansive government access to personal data is permitted

> DATA LOCATION

> NO LEGAL SAFEGUARDS

> SURVEILLANCE ENVIRONMENT

CONSEQUENCES

In Feb 2025, South Korea suspended DeepSeek:
- Data transferred to China/US without user consent
- No transparency or protection framework
- App removed, data deleted, and compliance rebuilt

Reinstated only after adding opt-outs and policy changes.

CONCLUSIONS

▶ **Collision Course**
Powerful GPAI + EU IP & data law → inevitable clash (copyright, trade-secrets, GDPR)**.**

▶ **The Rule-Book Thickens**
AI Act · Data Act · DMA → non-compliance becomes **existential.**

▶ **Build an EU-Proof Stack Now**
- GDPR opt-outs & dataset transparency
- Patent / Freedom-to-Operate checks
- FRAND licensing & response plan

Europe doesn't reject transformative AI,
it rejects unaccountable AI.
The firms that embed transparency and diligence will keep their 200 M-user head-start — the rest will discover it's cheaper to comply than to exit.

BIBLIOGRAPHY

DeepSeek: The Chinese AI app that has the world talking, *BBC*

What to Know About DeepSeek and How It Is Upending A.I., *The New York Times*

Why DeepSeek Shouldn't Have Been a Surprise, *Harvard Business Review*

DeepSeek shakes up AI sector – and other digital tech stories you need to know, *World Economic Forum*

Is AI Distillation By DeepSeek IP Theft?, *Winston & Strawn LLP*

Is DeepSeek's intellectual property protection sufficient to counter potential risks?, *Managing IP*

IA: il Garante privacy chiede informazioni a DeepSeek. Possibile rischio per i dati di milioni di persone in Italia, Garante per la Protezione dei Dati Personali

COMUNICATO STAMPA - Intelligenza artificiale: il Garante privacy blocca DeepSeek, *Garante per la Protezione dei Dati Personali*

DeepSeek, privacy zero: comincia in Italia lo scontro di diritti, *Agenda Digitale EU*

DeepSeek: Legal Considerations for Enterprise Users, *Ropes & Gray*

EU regulators scrutinize DeepSeek for data privacy violations, *UserCentrics*

US mulls penalties to block DeepSeek from buying American technology, NYT reports, *Reuters*

DeepSeek (chatbot), *Wikipedia*

OpenAI has evidence that its models helped train China's DeepSeek, *The Verge*

Italy's regulator blocks Chinese AI app DeepSeek on data protection, *Reuters*

DeepSeek Ai Under EU Scrutiny, GDPREU

DeepSeek changes rules of AI's great game, Financial Times

THANKS FOR YOUR ATTENTION!