

Themes	Subthemes	Definition	Number
Legal terms	Legal uncertainty	Legal issues are not definitively decided one way or another until settled in court. Legal decisions cannot necessarily be generalized to apply to other situations until that specific situation has been brought to trial	4
	Shrinkwrap agreement	An agreement that a buyer agrees to when they remove the shrink wrap from a (usually software) product.	6
Dataset user holds primary responsibility	User holds responsibility	The dataset user almost always holds liability for license violations.	7
	The dataset user almost always holds liability for license violations.	Licenses can be "bare", meaning provided as-is and not necessarily shown to the user, or as part of a license agreement, which is a type of contract and usually contains more restrictive stipulations.	4
	Strict liability tort	Copyright is generally considered strict liability tort, so users can be liable for damages even if the infringement was accidental.	4
Conditional liability of hosting platforms	Copyright regulation	Anyone found distributing copyrighted material without a license is liable for civil and criminal consequences.	6
	Limited liability	OCILLA protects service providers from liability so long as they comply with takedown requests from copyright owners.	4
Criminal consequences	Legality framework of copyright	Copyright is a criminal issue, not just a civil one. Criminal suits are brought by the government, while civil suits are brought by private individuals.	7
	Wide range of consequences	Penalties can include up to major fines and long prison sentences, and stack for multiple offenses.	5
	Getting a conviction	Prosecutors must prove willful infringement, which means knowing and intentionally malicious. This is very hard to prove, and so criminal cases are rarely brought.	5
	Criminal enforcement	Enforcing criminal cases is easier than civil ones because of dedicated law enforcement agencies, but international enforcement requires treaties and cooperation, and is reserved for extreme cases.	5
	Protected data	Licensing is not directly connected to regulation, but can insulate dataset owners from downstream misuse.	4
Civil consequences	Licenses to defray upstream liability	Licenses allow dataset owners to defray liability for copyright infringement of upstream data committed by downstream users. If a dataset is made up of copyrighted data but compliant with fair use,	5
	Difference between US and UK	Courts in the US generally find licenses binding without having to prove an agreement was established, whereas courts in the UK were more likely to demand that an agreement must have been established	3
	Harm and damages	When a party suffers material and demonstrable harm, they may file a civil suit seeking damages from the party that is responsible and/or liable for the harm.	7
Exceptions and exclusions	Invalid contracts	A contract that stipulates a party perform an illegal act or is not fully established with consent from both parties - a non-agreement that is not legally binding.	5
	Clickwrap agreement	An agreement that a user must click to accept, commonly when first using a software or digital service. Commonly accepted as binding.	4
	Browsewrap agreement	An agreement that the user must navigate to but is not forcibly shown or made to agree or disagree. Commonly accepted as non-binding.	4
	Fair use	Exceptional use cases for a copyrighted work that do not constitute infringement, including parody, education, the public good.	7
	Penalty clause	An clause parties may attempt to include in a contract demanding unreasonable damages in the event of a	4

		breach. These are commonly considered invalid and non-binding.	
	HIPAA	HIPAA does not apply to information brokers and data providers in the US; only insurance companies and medical professionals.	2
Real-world consequences	Difficult to enforce	A plaintiff must go through many long legal steps to get to trial, then through trial itself to get a judgement, then through many more steps to get enforcement of said judgement from a recalcitrant defendant.	6
	Expense of litigation	The legal process is unreasonably expensive for most normal individuals or smaller organization, precluding them from defending themselves or rendering enforcement against violations.	7
	Reasons to litigate	Due to the immense expense, plaintiffs are disincentivized from lawsuits unless the violator has money or other resources that can be seized as damages or the violation is causing severe harm.	6
	Corporate veil	Limited-liability companies protect ownership from liability incurred by the company. This adds additional difficulty to enforcement and can both protect bad actors from enforcement and protect dataset owners from bad actors.	3
	Future regulation	The US is unlikely to see regulation soon, while the EU will likely expand existing laws to cover AI-related copyright issues.	4
Limited awareness of license terms	Agreements are binding	Because agreements can be constructed rather than explicit, users may not realized they have entered into a legally binding contract when they actually have.	5
Ambiguity and complexity of licenses and licensing language	Length	Because licenses and agreements can be overly long, users rarely read them and may not understand exactly to what they are agreeing.	7
	Licenses vs contracts	Licenses, license agreements, and contracts are all slightly different but overlapping concepts and can be very confusing even for lawyers to understand.	6
	Guidelines are not binding	Usage guidelines may appear to be binding but actually are not. This could be intentional or accidental on behalf of the licensor - the researchers publishing the dataset may not understand the license as well as the user.	6
High costs and legal barriers to enforcement	Not worth trying to enforce	Since getting enforcement is so difficult and rarely rewarding, many owners do not bother to pursue enforcement. This lowers the risk of violating and increases the risk of illegitimate copyright trolls.	5
	International enforcement	International civil enforcement is almost always impossible, even in countries with strong rule of law and close relationships. There is very little incentive for international law enforcement to pursue civil enforcement.	4
	Trial in absentia	Allows a court to try a defendant in a lawsuit without them actually present. This is rarely used, and even when it is it does not solve the challenges of enforcement, only streamlines the trial process.	3
Imbalanced and unfair licensing structures	Adhesion contracts	One-sided contracts that are generally provided as-is. These contracts are often weighted in favor of the larger party, so courts are supposed to interpret the contract in a way most favorable to the less advantaged party. This is not well-enforced, however.	4
	Legal intimidation	Parties with larger amounts of money or lawyers on staff may use intentionally vague language to coerce licensees into agreeing to invalid terms the licensor knows they cannot afford to legally contest.	5