

Copyright and Fair Use

From APIs to AI Training Data

Security Course

Today's Agenda

1. Copyright Basics
2. The Fair Use Doctrine
3. Google LLC v. Oracle America Case
4. Transformative Use and Innovation
5. Copyright Detection Systems
6. Balancing Protection and Innovation

What is Copyright?

Purpose:

- Incentivize creativity and innovation
- "Promote the Progress of Science and useful Arts" (U.S. Constitution)
- Balance creators' rights with public access

What's Protected:

- Original works of authorship
- Literary, musical, dramatic, artistic works
- Software code (since 1980)

What's NOT Protected:

- Ideas, concepts, facts

Copyright Protections

Exclusive rights granted to copyright holders:

1. Reproduction (making copies)
2. Distribution (selling, sharing)
3. Derivative works (adaptations, modifications)
4. Public performance
5. Public display

Duration: Life of author + 70 years (for individuals)

Infringement: Unauthorized use of any exclusive right

The Fair Use Doctrine

Key concept: Not all unauthorized use is infringement

Fair use allows limited use of copyrighted material without permission for purposes like:

- Criticism and commentary
- News reporting
- Teaching and education
- Research and scholarship
- Parody

Question: How do we determine what's "fair"?

The Four-Factor Fair Use Test

Courts evaluate fair use based on four factors:

- 1. Purpose and character of use**
- 2. Nature of the copyrighted work**
- 3. Amount and substantiality used**
- 4. Effect on market value**

None of these factors is dispositive—courts weigh all four together

Factor 1: Purpose and Character

Asks: How and why is the work being used?

Favors fair use:

- Transformative use (adds new meaning/purpose)
- Educational, research, non-profit use
- Commentary, criticism, parody

Disfavors fair use:

- Commercial purposes
- Verbatim copying
- Superseding the original

Key concept: Transformativeness is often the most important consideration

Factor 2: Nature of the Work

Asks: What kind of work was copied?

Favors fair use:

- Factual works
- Published works
- Functional works (like software)

Disfavors fair use:

- Highly creative works (art, novels, music)
- Unpublished works

Rationale: Factual/functional works are closer to unprotectable ideas

Factor 3: Amount and Substantiality

Asks: How much was copied?

Considers:

- Quantitative amount (percentage copied)
- Qualitative significance (was the "heart" of the work taken?)

Not dispositive: Sometimes copying the entire work can be fair use

- Example: Thumbnail images in search results

Context matters: What was necessary for the transformative purpose?

Factor 4: Effect on Market Value

Asks: Does the use harm the copyright holder's market?

Considers:

- Does it substitute for the original?
- Does it harm potential licensing revenue?
- Does it harm the market for derivative works?

Not just actual harm: Also considers potential future markets

Case Study: Google LLC v. Oracle America

Background:

- Sun Microsystems created Java programming language
- Oracle acquired Sun in 2010
- Google created Android OS using Java APIs

The Dispute:

- Google copied ~11,500 lines of Java API declaring code
- Oracle sued for copyright infringement (\$9 billion in damages)
- Case went to Supreme Court (2021)

What's an API?

API = Application Programming Interface

Declaring code: Function signatures, method names, parameters

```
public class Math {  
    public static int max(int a, int b)  
}
```

Implementing code: The actual instructions that perform the task

```
public static int max(int a, int b) {  
    return (a > b) ? a : b;  
}
```

Google's position: Declaring code is not copyrightable (like a filing system)

Oracle's position: Declaring code is creative expression, protected by copyright

The Legal Journey

District Court (2012): APIs not copyrightable

Federal Circuit (2014): APIs ARE copyrightable (*reversed*)

Supreme Court (2018): Declined to review copyrightability

District Court (2016): Google's use was fair use

Federal Circuit (2018): NOT fair use (*reversed again*)

Supreme Court (2021): Google's use WAS fair use (*final decision*)

Key: Supreme Court assumed APIs were copyrightable but focused on fair use

Supreme Court's Analysis: Factor 1

Purpose and Character - Most Important Factor

Court found Google's use was **highly transformative**:

- Different purpose: smartphones vs. desktop computers
- Different context: mobile platform vs. server environment
- New users: mobile developers
- Enabled creation of new applications

Quote from decision:

"Google's use of the API was consistent with that creative 'progress' that is the basic constitutional objective of copyright itself."

Supreme Court's Analysis: Factor 2

Nature of the Work

Court found this favored Google:

- APIs are functional, not purely creative
- Declaring code acts like "organizational system"
- More like interface than artistic expression
- Designed for re-implementation

Key insight: The declaring code "is inextricably bound together with a general system...the division between copyrightable and uncopyrightable is harder to draw"

Supreme Court's Analysis: Factor 3

Amount and Substantiality

Court found Google's copying was appropriate:

- Only copied 0.4% of total Java API (11,500 of 2.86 million lines)
- Copied what was necessary for compatibility
- Reimplemented all functionality (wrote own implementing code)
- Needed to copy declaring code for developers to use existing skills

Principle: Amount copied must be viewed in context of transformative purpose

Supreme Court's Analysis: Factor 4

Effect on Market

Court found minimal harm to Oracle's market:

- Google and Oracle served different markets (mobile vs. server/desktop)
- Oracle had not entered smartphone market when Android launched
- Oracle's main concern was licensing fees, not market substitution
- Blocking Google would harm public interest in innovation

Decision: Google's use was fair use (6-2 decision)

Why This Case Matters

For software development:

- Protects API compatibility and interoperability
- Allows developers to use familiar interfaces
- Encourages innovation and new platforms
- Limits copyright holder control over functional elements

For innovation:

- Transformative use is a strong fair use defense
- Context and purpose matter more than amount copied
- Public interest in progress weighs heavily

For future cases: Provides framework for evaluating software and AI training

From Code to Content: Copyright Detection

Modern challenge: Millions of uploads daily on YouTube, TikTok, Instagram

Solution: Automated copyright detection systems

- YouTube Content ID
- Facebook Rights Manager
- TikTok copyright detection
- Spotify audio matching

How they work: Audio/video fingerprinting and matching

How Content ID Works

1. Rights holders upload reference files

- Original songs, videos, films

2. System creates digital "fingerprints"

- Audio: spectral analysis, waveforms, patterns
- Video: perceptual hashing, frame comparison

3. User uploads are scanned

- Compared against fingerprint database
- Matches flagged automatically

4. Rights holder chooses action

- Block the content

The Over-Blocking Problem

Challenge: Automated systems can't evaluate fair use

Examples of false positives:

- Film criticism with short clips (commentary/criticism)
- Parody music videos (transformative)
- Live performances of public domain works
- Background music in personal videos

Consequences:

- Content removed before human review
- Creators lose revenue
- Burden on users to dispute claims

The Appeals Process

YouTube's process:

1. Upload flagged by Content ID
2. Creator can dispute claim
3. Rights holder reviews dispute (30 days)
4. If rejected, creator can appeal
5. Rights holder can issue DMCA takedown
6. Creator can counter-notify
7. Rights holder must sue or content restored (10-14 days)

Problems:

- Power imbalance (rights holders control first review)
- Slow process

False Positives in the Wild

Real examples:

- NASA's Mars landing video flagged for copyrighted content
- Musician's own song flagged as copyright violation
- Game streamers flagged for in-game music
- Teachers' educational videos blocked
- Public domain classical music flagged

Fundamental issue: Algorithms can't understand context, purpose, or transformativeness

The Fair Use Dilemma

Rights holder perspective:

- Scale requires automation
- Manual review too expensive
- Risk of infringement without detection
- Lose revenue from unauthorized use

Creator perspective:

- Fair use is a legal right
- Automation prevents legitimate expression
- No compensation for false claims
- Disproportionate power imbalance

Balancing Copyright and Innovation

Too much protection (tech companies/public argue):

- Stifles innovation and creativity
- Limits access to knowledge
- Reduces interoperability
- Chills legitimate fair use

Too little protection (creators argue):

- Reduces incentive to create
- Business model disruption
- Enables unauthorized use
- Market value questions

Technical Solutions?

Possible improvements:

- Context-aware algorithms (detect commentary, education)
- Fair use filters (short clips, transformative indicators)
- Transparent scoring systems (show why content was flagged)
- Easier dispute processes
- Penalties for false claims

Limitations:

- Fair use is inherently context-dependent
- Requires human judgment
- No perfect algorithm for "transformativeness"

Policy Solutions?

Potential reforms:

- Require human review before monetization/blocking
- Faster appeal processes with neutral arbiters
- Penalties for abusive claims
- Safe harbors for educational/commentary use
- Transparency requirements (publish false positive rates)

DMCA reform proposals:

- Counter-notice reforms
- Bad faith claim penalties
- Fee-shifting for frivolous claims

Activity: Fair Use Analysis

Scenario: A YouTuber creates a 15-minute film criticism video including:

- 2-minute clip from the movie
- Frame-by-frame analysis with commentary
- Comparison to other films (short clips)

Your task: Apply the four-factor test. Is this fair use?

1. Purpose/character?
2. Nature of work?
3. Amount used?
4. Market effect?

Pair-share (5 minutes)

Copyright and AI: A Preview

Emerging question: Is training AI on copyrighted works fair use?

Arguments for fair use (AI companies):

- Transformative purpose (creating new outputs)
- Statistical pattern learning, not reproduction
- Promotes innovation and progress
- Like Google v. Oracle (functional use of existing works)

Arguments against (rights holders):

- Verbatim reproduction possible (memorization)
- Commercial purpose
- Substitutes for original works

Key Takeaways

1. **Copyright balances** creator incentives with public access
2. **Fair use doctrine** allows unauthorized use when transformative
3. **Four factors** evaluated together (purpose, nature, amount, market)
4. **Google v. Oracle** shows transformative use can justify extensive copying
5. **Automated systems** struggle with fair use's contextual judgment
6. **Over-blocking** is a real problem with significant consequences
7. **Balance is hard** but essential for innovation

Questions for Discussion

1. Should platforms be liable for user copyright infringement? Why or why not?
2. How can we design better copyright detection systems that respect fair use?
3. Does the Google v. Oracle decision apply to AI training on copyrighted data?
4. What role should human judgment play in copyright enforcement?

Next Lecture Preview

AI and Copyright

- Training AI models on copyrighted data
- Is it transformative use?
- The New York Times v. OpenAI lawsuit
- Authors Guild lawsuits
- Policy implications for AI development

Reading: Assigned case materials (check Canvas)

Thank You

Questions?

Office hours: [Your schedule]

Email: [Your email]

Assignment reminder: Copyright Detection Systems assignment on Canvas