

Politecnico di Torino, 29 June 2019



# *Copyright, AI and concentration of power*

Maurizio Borghi

Professor of Law

Centre for Intellectual Property Policy & Management

Bournemouth University

# Summary

- Why copyright matters (for AI)
- How copyright can foster monopolies *viz.* concentration of power (in the AI world)
- How can we save the (AI) “world”

# Some basic concepts

- Copyright (and relate rights)
- Public domain
- Fair use
- Exceptions and limitations

# Some more basics

- “Open access”
- Open licenses: GPL, Creative Commons

# The problem with copyright and AI

- Copyright is an “automatic” right
- Tech companies require repeated, systematic access to large amounts of data to train AI systems
- Good training data is crucial for creating accurate AI systems
- Some of the most valuable data are human-created works
- “Human” reading (viewing, listening) vs AI “reading”

# Examples

- Natural language processing – word embedding (word2vec)
- Machine translation
- Plagiarism detection
- Face recognition, voice recognition, etc...

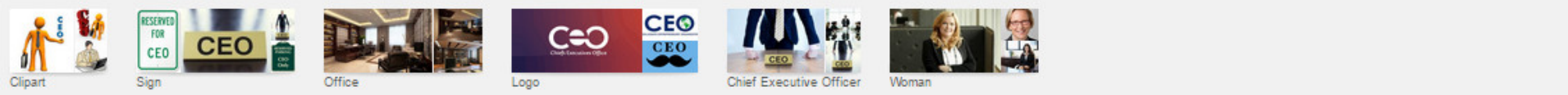
# How to feed the machine (with human-created works)

- Either you find a way to elude copyright, or...
- There are essentially two ways to secure access to a large corpus of © works: “buy-it” or “build-it”
- Buy-it approach: *e.g.* IBM’s Oncology Partnership, Google News, Google’s Patent Translate, ....
- Build-it approach: Facebook (and all social media)

# Buy-it or build-it, but in both ways...

- Poorly accountable (and possibly biased) AI systems
- The “black-box” phenomenon
- “Propertization” of datasets and concentration of power
- Not only tech giants: the case of iParadigms





# The third way

- Permitted & permissible uses
- Examples: use of “open access” repositories, CC, Wikipedia,...
- Public bodies are obliged (by law) or encouraged (by policy) to make data/works created by them available for AI research
- The problem with the third way

# The problem with the third way

- Risk-limiting solutions are always sub-optimal
- Quality of the data and possible bias due to use of “biased, low-friction data” (BLFD)
- Examples: Wikipedia, Open access journals, works in public domain – not always the best datasets to train AI systems

# Where are we now?

- Large / dominant players owning both algorithms' code and datasets to feed them
- Entrants in the AI market left with two options: either use BLFD or become “vassal” of tech giants
- But (you may say) copyright law has the solution: fair use!

# The fair use solution (and why it doesn't work)



- Not as “democratic” as it may seem: compare Google and Tixdaq
- But even if it were 100% predictable...
- The problem with “fair use exclusivity”, *i.e.* when the first fair-user takes the whole benefit of the permitted use and precludes second-comers from availing of the same benefit.
- Examples?

# Google Books

- Started in 2004 as “Google Print” (then Google Book Search)
- Scanning and making available for search “all world’s books” (now featuring 30M + books in 45 languages), in partnership with libraries
- Sued in 2005 by the Authors Guild; \$125M settlement agreement reached in 2009, but rejected by the court
- Finally decided as fair use (*AG v Google*; *AG v HathiTrust*)

# Google Books



Google

**Web** Images Videos Shopping News More ▾ Search tools

About 1 results (0.49 seconds)

Cookies help us deliver our services. By using our services, you agree to our use of cookies.

[Learn more](#)

## Stand Still Like the Hummingbird - Page 159 - Google Books ...

[books.google.co.uk/books?isbn=0811203220](https://books.google.co.uk/books?isbn=0811203220)

Henry Miller - 1962 - Fiction

But perhaps the greatest thing to be gained from the reading of books is the desire to truly communicate with one's fellow man. To read a book properly is to ...

# Google Books

**GET PRINT BOOK**

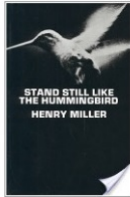
No eBook available

New Directions Publishing  
Amazon.co.uk  
Book Depository - £9.99  
Waterstone's  
WHSmith  
Blackwell

---

Find in a library  
Find local bookshops  
All sellers >

## Stand Still Like the Hummingbird



Henry Miller  
New Directions Publishing, 1962 - Fiction - 194 pages  
★★★★★  
36 Reviews

One of Henry Miller's most luminous statements of his personal philosophy of life, *Stand Still Like the Hummingbird*, provides a symbolic title for this collection of stories and essays.

[Preview this book >](#)

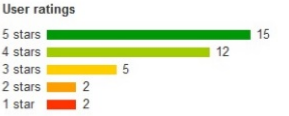
**New! Shop for Books on Google Play**

Browse the world's largest eBookstore and start reading today on the web, tablet, phone, or e-reader.

[Go to Google Play Now >](#)

### What people are saying - [Write a review](#)

- Review: Stand Still Like the Hummingbird**  
User Review ★★★★★ - Tara Bree - Goodreads  
I learned that I am not really into Henry Miller. It is difficult for me to concentrate on what he is writing. Perhaps it's because I am coming out of 7 years of schooling and am tired of extreme ... [Read full review](#)
- Review: Stand Still Like the Hummingbird**  
User Review ★★★★★ - Matt Macneil - Goodreads  
complex and confident [Read full review](#)



- My library
- My History
- Books on Google Play

### Related books





# Google Books



WEB IMAGES VIDEOS MAPS NEWS MORE



"perhaps the greatest thing to be gained from the reading of books i | 🔍

71.900 RESULTS Narrow by language ▼ Narrow by region ▼

No results found for "perhaps the greatest thing to be gained from the reading of books is the desire to truly communicate with one's fellow man".

Showing results for perhaps the greatest thing to be gained from the reading of books is the desire to truly communicate with one's fellow man (without quotes).

Answers



"perhaps the greatest thing to be gained from the reading of books is the desire to t

Find Answers

The Web  
 UK Only

Everything ▶

Your search for "perhaps the greatest thing to be gained from the reading of books is the desire to truly communicate with one's fellow man" did not match with any Answers results.

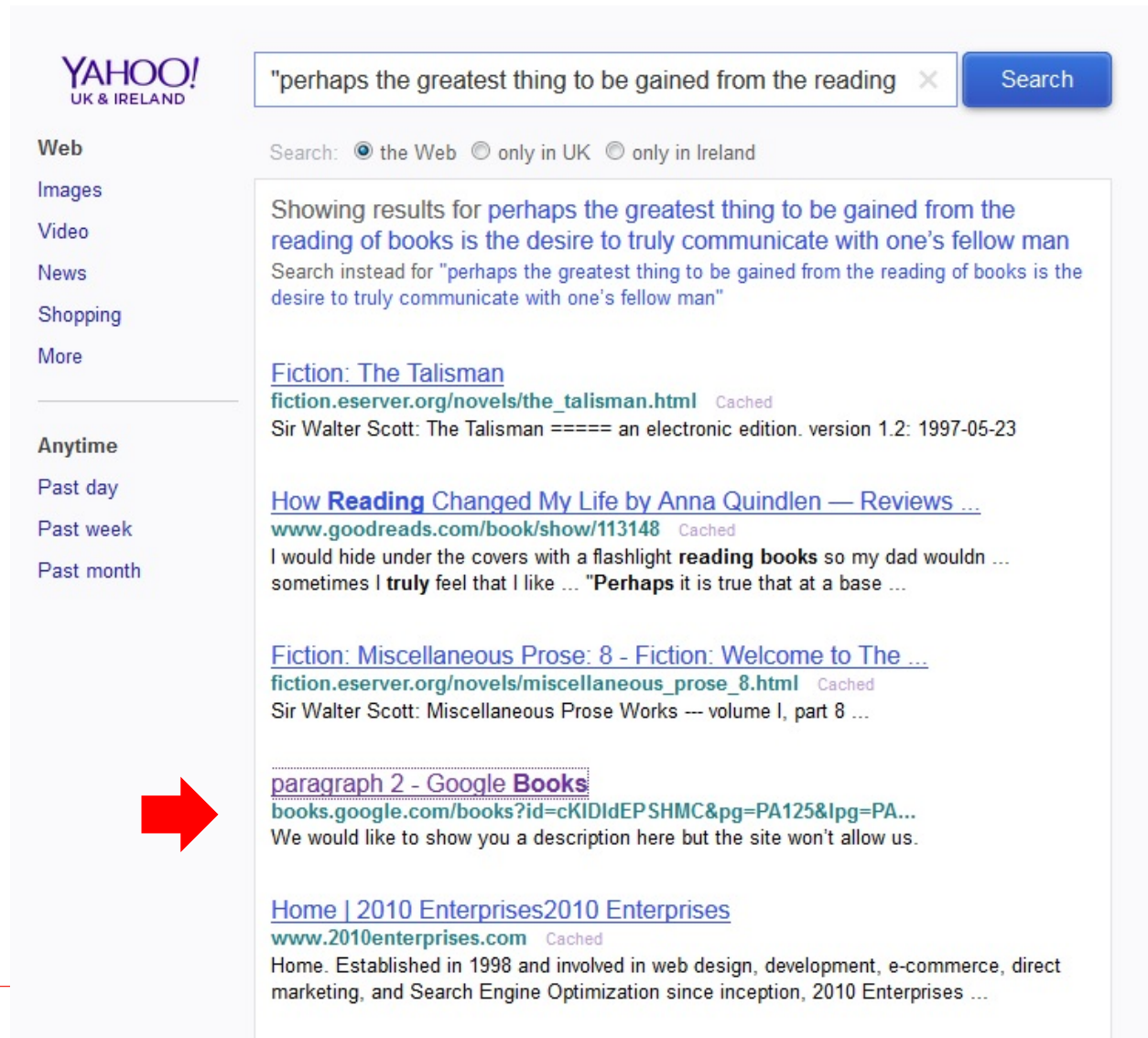
Images

Videos

Search Suggestions:

- Make sure all words are spelled correctly
- Try different keywords
- Try more general keywords
- Try fewer keywords
- View more [Search Tips](#)

# Google Books



The screenshot shows a Yahoo! search interface. The search bar contains the text "perhaps the greatest thing to be gained from the reading" and a search button. Below the search bar, there are radio buttons for "the Web", "only in UK", and "only in Ireland". The search results are displayed in a list format. The first result is a link to "Fiction: The Talisman" from fiction.eserver.org. The second result is a link to "How Reading Changed My Life by Anna Quindlen — Reviews ..." from www.goodreads.com. The third result is a link to "Fiction: Miscellaneous Prose: 8 - Fiction: Welcome to The ..." from fiction.eserver.org. The fourth result is a link to "paragraph 2 - Google Books" from books.google.com, which is highlighted with a red arrow. The fifth result is a link to "Home | 2010 Enterprises2010 Enterprises" from www.2010enterprises.com.

YAHOO!  
UK & IRELAND

Search: "perhaps the greatest thing to be gained from the reading" Search

Search:  the Web  only in UK  only in Ireland

Showing results for perhaps the greatest thing to be gained from the reading of books is the desire to truly communicate with one's fellow man  
Search instead for "perhaps the greatest thing to be gained from the reading of books is the desire to truly communicate with one's fellow man"

[Fiction: The Talisman](#)  
[fiction.eserver.org/novels/the\\_talisman.html](http://fiction.eserver.org/novels/the_talisman.html) Cached  
Sir Walter Scott: The Talisman ===== an electronic edition. version 1.2: 1997-05-23

[How Reading Changed My Life by Anna Quindlen — Reviews ...](#)  
[www.goodreads.com/book/show/113148](http://www.goodreads.com/book/show/113148) Cached  
I would hide under the covers with a flashlight **reading books** so my dad wouldn't ...  
sometimes I **truly** feel that I like ... "**Perhaps** it is true that at a base ...

[Fiction: Miscellaneous Prose: 8 - Fiction: Welcome to The ...](#)  
[fiction.eserver.org/novels/miscellaneous\\_prose\\_8.html](http://fiction.eserver.org/novels/miscellaneous_prose_8.html) Cached  
Sir Walter Scott: Miscellaneous Prose Works --- volume I, part 8 ...

[paragraph 2 - Google Books](#)  
[books.google.com/books?id=cKIDIdEP SHMC&pg=PA125&lpg=PA...](http://books.google.com/books?id=cKIDIdEP SHMC&pg=PA125&lpg=PA...)  
We would like to show you a description here but the site won't allow us.

[Home | 2010 Enterprises2010 Enterprises](#)  
[www.2010enterprises.com](http://www.2010enterprises.com) Cached  
Home. Established in 1998 and involved in web design, development, e-commerce, direct marketing, and Search Engine Optimization since inception, 2010 Enterprises ...

# Turnitin

- Students' coursework submission → *wholesale copying and storage of copyright-protected content for plagiarism detection* (fair use in *AV v iParadigm*, 2009)
- Text mining from web pages (45+ billion), academic publications (130+ million) and papers previously submitted by students for marking (1 billion!)
- AI optimization + development of related services: online marking, “authorship investigation”, automated marking (?)

# AI monopolies

- Not just monopolies on the exploitation of digital objects (*e.g.* digital works), but monopolies over (certain) uses made possible by AI
- Mainly operating in two-sided markets
- Mainly non-replicable uses
- Similar (but not akin) natural monopolies

# Unlocking AI monopolies



- Antitrust measures: doctrine of essential facilities (aka ‘refusal to license’ – ECJ *Magill*, C-241 – 242/91P) – *too slow or unavailable to remedy IP abuses*
- D.I.Y. measures – *e.g.* Internet Archive hosting  $\approx$  1M public domain books downloaded from Google Books and uploaded on their servers by users
- Use copyright as a basis to create “AI Commons”. But how?

# Towards AI Commons

- Legal basis: DSM Copyright Directive, art. 4: *mandatory exception for text and data mining*

# Art. 4 DSM Copyright Directive

1. Member States shall provide for an exception or limitation to the rights [*of reproduction of works and extraction of database*] for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining.
2. [...]
3. The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine readable means in the case of content made publicly available online.