



E-Discovery e Intelligenza Artificiale

DFA Open Day 2018

Andrea Tognolini, 5 luglio 2018
Università degli Studi di Milano



The better the question. The better the answer.
The better the world works.

Agenda

1

E-Discovery

Electronically Stored Information

Processo

Scenari critici

2

Technology-Assisted Review

Apprendimento e classificazione

Vantaggi

Adozione in casi internazionali

Questioni aperte

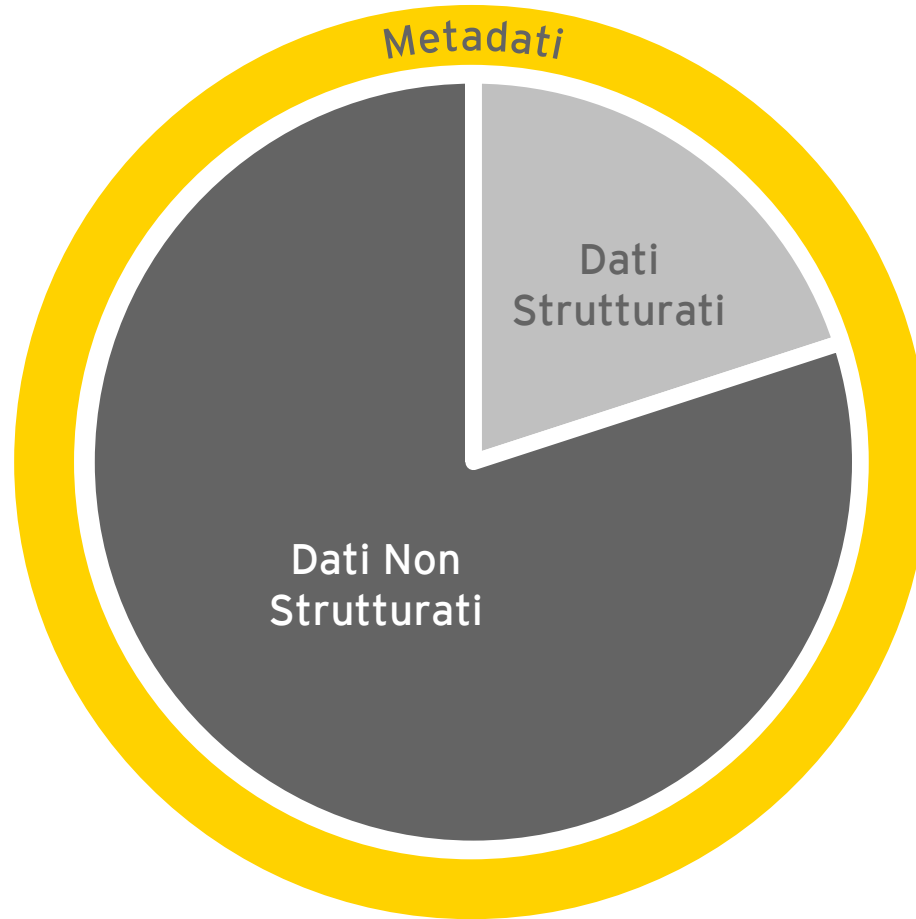
1. E-Discovery



E-Discovery is the process of ① identifying, ② preserving, ③ collecting, ④ processing, ⑤ searching, ⑥ reviewing, and ⑦ producing Electronically Stored Information (ESI) that may be relevant to a civil, criminal, or regulatory matter.

(Federal Courts Law Review, 2013. Grossman-Cormack glossary of technology-assisted review)

Electronically Stored Information



Metadati

- ▶ Autore
- ▶ Data di creazione
- ▶ Data di modifica

Dati Strutturati

- ▶ CRM
- ▶ Database
- ▶ Transaction System

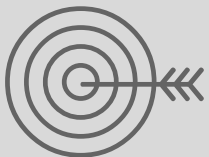
Dati Non Strutturati

- ▶ Text, Spreadsheet, Presentations
- ▶ Email
- ▶ Graphics

Processo E-Discovery

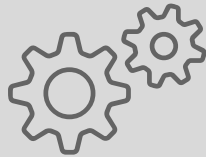
- ▶ Individuare i **soggetti** coinvolti e il materiale digitale in loro possesso
- ▶ Raccogliere il **materiale digitale**, con procedure che garantiscano l'integrità e la confidenzialità dei dati

1



Identificazione e Raccolta

2



Elaborazione

- ▶ Processare i dati acquisiti, per isolare il materiale personale, escludere duplicati, recuperare il materiale cancellato

- ▶ Definire **keyword** da applicare alla base documentale
- ▶ **Review manuale** dei documenti individuati
- ▶ Ottimizzare le keyword se necessario, per ridurre il numero di falsi positivi



3

Analisi e Revisione



4

Reporting

- ▶ Quando tutto il materiale rilevante è stato rivisto, preparare **report** o **dashboard**, secondo le necessità

Scenari critici del processo tradizionale

Criticità

- ▶ Grande numero di **documenti**
- ▶ **Tempo** limitato per completare la revisione
- ▶ Numero ristretto di revisori (400 elementi/persona/giorno)

Soluzione

Individuare **cluster** di documenti in base al contenuto, per permettere la revisione massiva

Criticità

- ▶ È necessario individuare velocemente i documenti potenzialmente rilevanti per il caso
- ▶ Il tempo e le risorse sono limitate

Soluzione

Il sistema calcola autonomamente la potenziale rilevanza di ciascun documento nel caso, sulla base di esempi

Criticità

- ▶ È necessario produrre una base documentale per la controparte in una disputa
- ▶ Tutto il materiale confidenziale e riservato non può essere condiviso

Soluzione

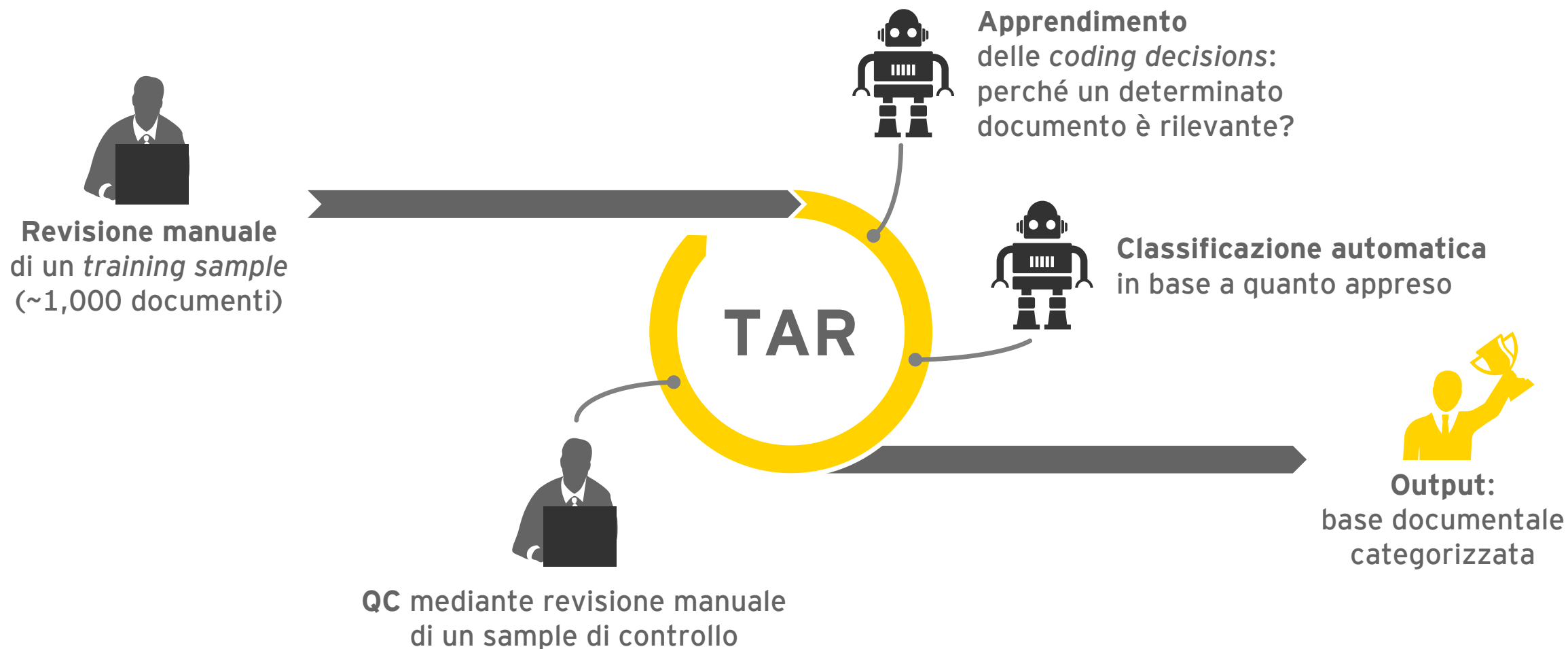
Trovare tutti documenti dal contenuto **simile** a un gruppo di esempio già individuato

Per tutto questo, abbiamo bisogno che la «macchina» possa comprendere il contenuto dei documenti sotto esame

2. Technology-Assisted Review



Technology-Assisted Review (TAR)



TAR: Apprendimento e classificazione

LSI

Latent Semantic Indexing

- ▶ Analizza i termini contenuti nei documenti sotto analisi
- ▶ Crea un insieme di concetti
- ▶ Si basa sul principio che parole di significato simile compaiono in contesti simili (ie, vicine ad altre parole di significato simile)



SVM

Support Vector Machines

- ▶ Sulla base di esempi, apprende quali concetti rendono un documento potenzialmente rilevante
- ▶ Calcola la probabilità un documento non ancora rivisto di essere rilevante per il caso, valutando la sua somiglianza rispetto al gruppo di esempio

TAR: Vantaggi

Dare **priorità** alla revisione di determinati documenti

Fare **quality check** del processo di *mail review*

Ridurre il **volume** di materiale da revisionare manualmente



Superare il concetto di **keyword** per la selezione dei documenti

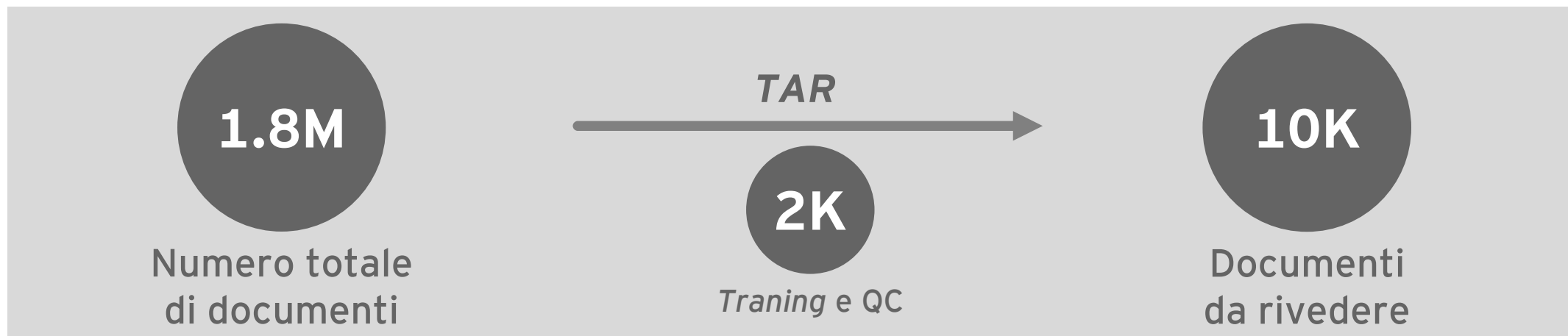
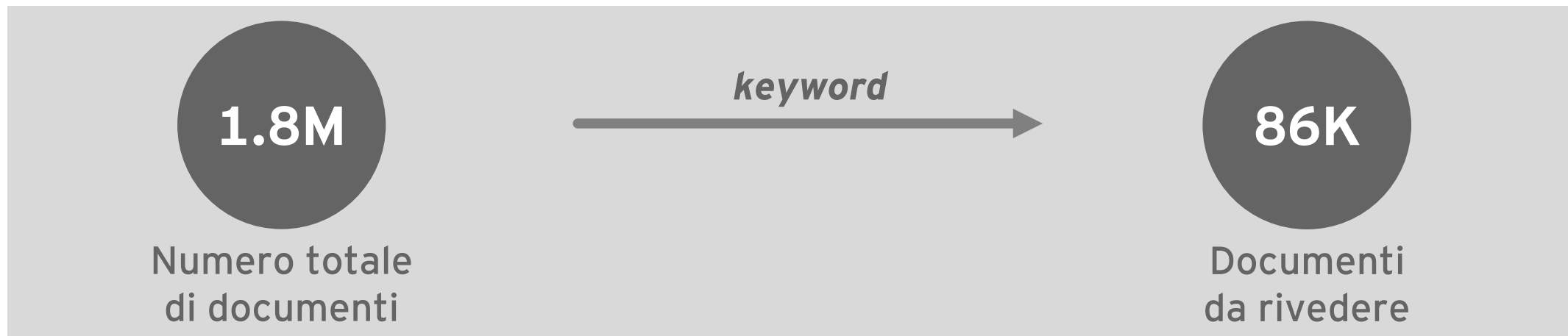
Identificare documenti molto simili tra loro (**near-duplicates**)

TAR: Adozione in casi internazionali

TAR viene ormai riconosciuto (e talvolta consigliato) come un metodo valido per la ricerca di materiale rilevante, al fine di ① ridurre i costi dovuti alla *document review*, ② rendere più efficace il processo di revisione, ③ accrescere la consistenza della revisione



TAR: Un caso d'uso



Source: <https://www.relativity.com/customers/inventus/>

TAR: Questioni aperte



La corte deve approvare esplicitamente l'utilizzo di TAR?
Può imporne l'utilizzo?



In quali situazioni è preferibile utilizzare TAR?



È preferibile o appropriato pre-filtrare la base documentale sulla base di alcune parole chiave, prima di applicare TAR?

Grazie
per l'attenzione

- ▶ *The Sedona Conference TAR Case Law Primer (2017).*
- ▶ Maura R. Grossman and Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review (2011).*
- ▶ Maura R. Grossman and Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review, with Foreword by John M. Facciola, U.S. Magistrate Judge (2013).*