Big Data, Artificial Intelligence and Privacy

Stephen Kai-yi Wong, Barrister
Privacy Commissioner for Personal Data, Hong Kong
Big Data and Artificial Intelligence
Big Data

• massive scale of collection, processing, combination and aggregation of data
Big Data Analytics

- transports us from past to future – make predictions and decisions
“Information is the oil of the 21st century, and analytics is the combustion engine.”

- Peter Sondergaard, Senior Vice President, Gartner Research.
Big Data Analytics

- beneficially used in areas of scientific and medical research, product development, education, marketing and building smart cities, etc.
Artificial Intelligence (AI)

- **machine learning** – algorithms which can learn and evolve without the need for human intervention

- **thrive with big data analytics** – discover patterns and correlations to make predictions and decisions
“Machines will be capable...of doing any work a man can do”
~ Herbert Simon, scientist
Applications of AI

- Rubik’s cube robot
- Voice-activated assistant
- Deep Text
- AI Judge
Privacy Implications of Big Data and AI
6 保障資料原則

Data Protection Principles

1 收集目的及方式 Collection Purpose & Means

資料使用者必須依合法和公平的方式，收集個人的個人資料，其目的應直接與其職能或活動有關。

Personal data must be collected in a lawful and fair way, for a purpose directly related to a function/activity of the data user.

All practicable steps shall be taken to notify the data subjects of the purpose of data collection, and the classes of persons to whom the data may be transferred.

Data collected should be necessary but not excessive.

2 準確性、儲存及保留 Accuracy & Retention

資料使用者須採取切實可行的步驟以確保持有的個人資料準確無誤，而資料的保留時間不應超過達致原來目的的實際所需。

Practicable steps shall be taken to ensure personal data is accurate and not kept longer than is necessary to fulfil the purpose for which it is used.

3 使用 Use

個人資料只能用於收集時指定的目的或直接相關的目的，除非得到資料當事人自願和明確的同意。

Personal data is used for the purpose for which the data is collected or for a directly related purpose, unless voluntary and explicit consent is obtained from the data subject.

4 保安措施 Security

資料使用者須採取切實可行的步驟，保障個人資料不會經由無權或意外地被查閱、處理、刪除、喪失或使用。

A data user needs to take practical steps to safeguard personal data from unauthorised or accidental access, processing, erasure, loss or use.

5 透明度 Openness

資料使用者須採取切實可行的步驟來公開其處理個人資料的政策和行事方式，並交代其持有的個人資料類別和用途。

A data user must take practicable steps to make personal data policies and practices known to the public regarding the types of personal data it holds and how the data is used.

6 查閱及更正 Data Access & Correction

資料當事人有權要求查閱其個人資料；若發現有關個人資料不準確，有權要求更正。

A data subject must be given access to his personal data and to make corrections where the data is inaccurate.
• massive, ubiquitous and invisible data collection

• how do data subjects know when and what kinds of data is collected?

• how do data subjects know if the data collected is excessive or not?

• Principle 1 on data collection
Challenge to Notice

- **Principle 1** notification requirement

- how to provide **meaningful notice** to data subjects and prevent “notice fatigue”?
Challenge to Notice

Big data “thrives on surprising correlation and produces inferences and predictions that defy human understanding… how can you provide notice about the unpredictable and unexplainable?”

~ Paul Ohm, Professor of Law at the Georgetown University Law Center
Algorithmic Transparency

- Big data and AI use algorithms to find correlations and make predictions from data.
- How to discover the logic or rationale behind the decisions made by big data or AI?
- Principle 5 on transparency.
• can data subjects ascertain whether the data collected by AI is true, fair or accurate?

• can data subjects object to unfair or biased decisions made from inaccurate data?

• Principle 2 on data accuracy

• Principle 6 on data correction
Perpetuate Bias – Real Life Examples

- **big data analytics** - because a number of customers of a shop had poor credit card repayment records, a credit card company rated another customer of that shop as higher credit risk

- in a **beauty contest**, AI was used to pick the winners, eventually no winners were ethnic minorities
Perpetuate Bias – Real Life Examples

• big data analytics predicts that those online job-seekers who use deliberately installed Internet browsers will outperform other job-seekers who use the factory-default browser of the computer

• a company used AI to identify terrorists, and claimed that the AI is also able to identify everything from great poker players, extroverts, geniuses to white collar-criminals
Unexpected Data Use

- data may be combined, aggregated and used in unexpected ways
- Principle 3 on data use limitations
Public Domain Data

• data may be collected from public domain for big data and AI analytics

• Principle 3 on data use limitations

• respect individuals’ reasonable expectations
Risk of Re-identification

- AI report of UK Government Office for Science: growth in AI and big data analysis facilitate re-identification of de-identified personal data

- infer private information from public data

- manage risk of re-identification

- Principle 4 on data security
EU General Data Protection Regulation:

- Article 4: defines “profiling” as automated processing of personal data and using data to evaluate personal aspects
- Article 13: transparency – disclose to individuals the profiling, logic and envisaged consequences
- Article 22: right to avoid being subject to a decision if based solely on profiling and produces legal effects
• who should be responsible for the act or decision made by big data or AI analytics?

• to what extent should a human being be responsible?

• would privacy law apply if personal data is collected and processed by AI machines alone?
Protect, Respect Personal Data

- respect the dignity and rights of human beings
- maintain high level of transparency
- **Privacy by Design**: adopt privacy-friendly approach in design phase
- **Privacy Management Programme**: from compliance to accountability
Protect, Respect Personal Data

• Privacy Management Programme: A Best Practice Guide

Protect, Respect Personal Data

- Isaac Asimov’s “The Three Laws of Robotics”:
  - 1st law: not to injure humans
  - 2nd law: obey humans’ orders except conflict with 1st law
  - 3rd law: protect own existence except conflict with 1st or 2nd law

What are your rules?
Other Topical Issues
Know Your Customer (KYC)

- Law firms may collect, use and retain lots of personal data for KYC.
- **Principle 1** on data collection (fair & not excessive).
- **Principle 3** on data use limitation (NB: apply to data in public domain).
- **Principle 2** on data retention (no longer than necessary).
  - any statutory or regulatory obligations? exceptional circumstances?
Know Your Customer

- Hong Kong Monetary Authority: banks should adopt a risk-based approach
  - transparency – list information required and explain rationale
  - reasonableness – only collect relevant information; not “zero failure” regime
  - information obtained should be retained for six years after the end of business relationship
Keep Your Phone Number Safe

- phone number may be as important as your ID card number, and may reveal your identity

- “3 billion phone numbers and identities exposed by mobile apps, investigation finds”
  FactWire, 20 Nov 2016
Bring Your Own Device (BYOD)

- employees use their own mobile devices to access and work with their employers’ organisational information

- protective measures should respect private information on BYOD equipment

- establish BYOD policy, provide staff guidelines and apply technical solutions
Bring Your Own Device (BYOD) Information Leaflet

- Highlight privacy risks in developing BYOD policy
- Suggest BYOD best practices

PCPD conducted “Privacy Sweep” to examine how fitness bands communicate privacy-related matters to users.

- **Observations:**
  - Lack of transparency in privacy policies.
  - Request access to data in smartphones which may be unnecessary (e.g. location, social media accounts).
  - No clear guide to erase data.
  - Insufficient data security.
Advice to Users

- research on privacy impact before purchase
- use pseudonyms for account registration
- set up dedicated email accounts
- review default settings
- patch firmware and update app
- purge data before disposal/resale
Advice to Manufacturers

• provide clear privacy policies to users in simple language
• adopt “Privacy by Design”:  
  – minimise data collection;  
  – incorporate sufficient security safeguards;  
  – adopt privacy-friendly settings  
• offer opt-out option if access to data is unnecessary  
• provide clear instructions for data erasure
Thank You!

Stay tuned for updates on
www.privacyconference2017.org