КОМИСИЯ ЗА ЗАЩИТА
НА ЛИЧНИТЕ ДАННИ

ГОЛЕМИ БАЗИ ДАННИ (BIG DATA) И СВЪРЗАНАТА С ТЯХ ВЪЗМОЖНОСТ ЗА ПРОФИЛИРАНЕ

БРОШУРА ЗА АДМИНИСТРАТОРИТЕ
НА ЛИЧНИ ДАННИ

Настоящият информационен материал има за цел да подпомогне практическото прилагане на Регламент (ЕС) 2016/679 (Общ регламент относно защитата на данните), да даде разяснения на някои ключови въпроси при обработването на „Големи информационни масиви“ („Големи бази данни“, Big Data) и свързаната с тях възможност за профилиране. Същият няма задължителен характер и не претендира за изчерпателност.

Бързото технологично развитие, широкото използване на компютрите, интернет и цифровите технологии, глобализацията създават нови предизвикателства пред защитата на личните данни.

В ерата на цифровия свят, в който живеем днес, всяка дейност оставя цифрова следа, която може да бъде събрана, обработена и оценена или анализирана. С новите информационни и комуникационни технологии се събират, записват и анализират все повече данни.

Определение и видове

За първи път през 2001 г. Laney споменава трите V-та, които са в основата на определянето на „големите данни“: Volume (обем), Velocity (скоростта на нарастване) и Variety (многообразие).

„Големи информационни масиви“ („Големи бази данни“, Big Data) е модерен термин, който най-общо включва „нарастващата технологична способност за събиране, обработване и извличане на нова и прогнозна информация от голям обем, скорост и разнообразие от данни“.

Понятието „големи информационни масиви“ обхваща едновременно самите данни и анализа на данните. Според начина на представяне могат да се класифицират в три категории – структурирани, полуструктурирани и неструктурирани.

Източници на големи бази данни

Източниците на данните са най-различни по вид и включват:

- Хора и техни лични данни, електронни устройства, машини или датчици, информация за климата, спътникови изображения, цифрови снимки и видеоматериали или GPS сигнали;
- Данните, генерирани от хора чрез мобилни приложения, в интернет, през социалните мрежи или плащания, данни на публичната администрация;
- Интернет свързани устройства/интернет на предметите/ интернет на вещите (Internet of Things, IoT);
- Метаданните;
- Данни от обществените сфери.

Начини за обработка на големи бази данни

Системите за работа с големи данни се използват за извличане и разкриване на неочевидни зависимости, скрити корелации между на пръв поглед неизвестни и несвързани една с друга величини или за обобщаването им по нови начини, които са разбираеми и полезни за потребителите. За тази цел се прилагат интелигентен анализ на данни и текст (Data Mining и Text Mining), различни аналитични техники и методи за мрежов анализ, анализ на мултимедии и социални медии и др.

За обработване на големите бази данни се използват:

- Паралелни и разпределени парадигми (напр. облачните технологии);
- Изкуствения интелект (AI);
- Други подходи и инструменти (например NoSQL бази данни).

Ползи и рискове, свързани с големите информационни масиви

При отговорно боравене големите масиви от данни може да донесат значителни ползи и ефективност за обществото и отделния човек. Сериозно безпокойство обаче предизвиква действителното и потенциалното въздействие на обработката на огромни количества данни върху правата и свободите на хората, включително неприкосновеността на личния им живот. Затова предизвикателствата и рисковете, свързани с големите масиви от данни, налагат по-ефективна защита на данните.

Ползи при обработването на големи бази данни:

- Обхващане на големи масиви от данни;
- Бързо импортиране на нови данни;

- Обработване в реално време на информацията при кратки срокове;
- Възможност за множество и едновременни искания;
- Анализиране на различни видове информация (снимки, текстове или номера);
- Осигуряване на нови средства за профилиране и целева реклама;
- Разкриване на закономерности между различните източници и масиви от данни;
- Анализът на информацията в реално време може да бъде използван за подобряване на въведените системи;
- Нарастване на броя на трансграничните потоци от данни;
- Значителни и видими ползи в ежедневието: интернет търсачките улесняват достъпа до значителни обеми информация и знания; услугите за социални мрежи дават възможност на хората в целия свят да общуват, да изразяват мнението си и да мобилизират подкрепа за различни каузи, възползване от ефективни и ефикасни маркетингови техники, които стимулират икономиката;
- Технологиите и обработването на лични данни представляват незаменими инструменти за държавните органи в борбата им срещу престъпността и тероризма.

Рискове при обработването на големи бази данни:

- Физическите лица все по-често оставят лична информация, която е публично достъпна и в световен мащаб;
- Неправомерна употреба на големите информационни масиви от лица с достъп до масивите от информация посредством

манипулиране, дискриминация или потискане на отделни лица или на конкретни групи в обществото;

• По отношение на защитата на основните права като правото на неприкосновеност на личния живот и защитата и сигурността на данните;

• Във връзка със свободата на изразяване на мнение и недопускането на дискриминация;

• При обучаването на устройства с изкуствен интелект (AI), като невронни мрежи, и статистически модели с цел предвиждане на определени събития или поведение често данните за обучението са със съмнително качество и не са неутрални;

• Съвременните техники за обработка и анализ, вкл. без намесата на човека, предоставят безпрецедентен поглед върху човешкото поведение, личния живот и обществото;

• Чувствителна информация за лица може да се изведе и от нечувствителни данни;

• Технологичният напредък и възможностите за анализи на големи информационни масиви, изкуственият интелект (AI) и машинното самообучение улесниха създаването на профили и вземането на автоматизирани решения, които имат потенциал да окажат значително въздействие върху правата и свободите на физическите лица;

• Широкоразпространената достъпност на лични данни в интернет и от интернет свързани устройства/интернет на предметите/интернет на вещите, както и възможността да се установяват корелации и да се създават връзки, могат да позволят определяне, анализ и предвиждане на личностни или поведенчески аспекти на даден човек и на неговите интереси и навици (профилиране);

• Информацията за поведението и нагласите днес се използва за определяне на личността на съответното лице. Комбинираната информация за „харесванията“ в социалните мрежи, данните от проследяването, слушаната музика или гледаните филми дава възможност да се изгради ясна картина на личността на дадено лице, позволявайки на предприятията да публикуват целеви реклами и/или информация в съответствие с „личността“ на това лице;

• Техниките и новият софтуер за обработване оценяват в реално време информацията за това какво харесва дадено лице, какво разглежда то, когато пазарува онлайн, или какво слага в кошницата с покупки онлайн и могат да предложат „продукти“, които може да представляват интерес въз основа на събраната информация (профилиране);

• Възможно е процесите на профилирането и автоматизираното вземане на решения да не са видими. Физическите лица може да не знаят, че са подложени на профилиране, или да не разбират какво включва това.

Псевдонимизацията, анонимизиране или криптиране на личните данни, са подходящи предпазни мерки за намаляване на рисковете за съответните субекти на данни, когато личните данни се използват в приложения за големи информационни масиви. Договорните задължения следва да гарантират, че анонимизираните данни няма да бъдат повторно идентифицирани чрез използване на допълнителни съотношения посредством комбиниране на различни източници на данни. Употребата на криптиране от край до край следва да се насърчава и когато е необходимо да се прави задължително, в съответствие с принципа на защита на данните при проектирането (чл. 25 от Регламент (ЕС) 2016/679).

Задължения на администраторите на лични данни

Съгласно Регламент (ЕС) 2016/679, в контекста на обработването на големи бази данни, администраторите на лични данни са задължени да:

- Уведомят субектите на данни за съществуването на автоматизирано вземане на решения, включително профилиране *(член 12)*, като уведомлението съдържа и съществена информация относно използваната логика при профилирането и предвидените последствия от това обработване за лицата *(член 13, параграф 2, буква е)*;

- Предприемат подходящи мерки, за да се гарантират правата, свободите и законните интереси на субектите на данни. Това включва най-малко правото на човешка намеса от страна на администратора и възможността субектът на данните да изрази гледната си точка и да оспори решение, което се основава на автоматизирано обработване на личните му данни *(член 22, параграф 3)*;

- Изрично да предоставят подробности на вниманието на субекта на данни относно правото му на възражение *(член 21, параграфи 1 и 2)*, както и да ги представят по ясен начин и отделно от всяка друга информация *(член 21, параграф 4)*;

- Спазват принципа за свеждане на данните до минимум, както и изискванията за ограничение на целите и принципите за ограничение на съхранението *(член 5)*;

- Личните данни в контекста на профилирането следва да бъдат обработвани добросъвестно, законосъобразно, пропорционално и за конкретни и легитимни цели *(член 5)*;

- Предприемат мерки да коригират факторите, които водят до неточности в личните данни *(член 5)*;

• Въведат надеждни мерки за постоянна проверка и гарантиране, че повторно използваните или получените по косвен път данни са точни и актуални;

• Демонстрират, че субектите на данни разбират с какво точно се съгласяват, и също така да помнят, че съгласието невинаги представлява подходящо основание за обработване;

• При наличие на легитимен интерес, трябва да се извърши балансиране, за да се оцени дали пред интересите на администратора преимущество имат интересите или основните права и свободи на субекта на данните;

• Ограничат рисковете или грешките, които профилирането може да предизвика;

• Периодично да оценяват качеството на използваните данни и алгоритми;

• Гарантират, че профилирането и автоматизираното вземане на индивидуални решения (независимо дали включва профилиране или не) не се използват по начини, които оказват необосновано въздействие върху правата на физическите лица;

• Ако от профилирането се извличат чувствителни предпочитания и характеристики, администраторът следва да гарантира, че: 1. обработването не е несъвместимо с първоначалната цел, 2. е определил законосъобразно основание за обработването на специалните категории данни, както и 3. е информирал субекта на данни относно обработването;

• Извършват анализ на риска и оценка на въздействието върху защитата на данните;

• Определят видовете потребители в информационните системи;

• Утвърдят правила за отделните потребители на информационните системи, функционалните им задължения и процедурите за тяхната дейност, в които достатъчно ясно да са разписани принципите на взаимодействие на отделните потребители;

• Утвърдят правила за обработка на личните данни за всяка една от подържаните информационни системи;

• Утвърдят правила и процедури за защита на личните данни, в т. ч. информационната/киберсигурност;

• Утвърдят вътрешни правила за обучение и тренировка на служителите за действия в случаи на незаконосъобразно обработване на лични данни;

• Поддържат системни дневници (логове) на действията на потребителите на информационните системи;

• Имат налична документация къде и как на етапа на проектиране на информационните системи е приложен чл. 25 от Регламент (ЕС) 2016/679 (защита на данните на етапа на проектиране и по подразбиране);

• Утвърдят политики и процедури за защита на личните данни, които гарантират спазването на Регламент (ЕС) 2016/679, при наемане на облачни услуги и тяхното управление по отношение на достъп / преносимост / възстановяване / унищожаване на данни в т. ч. извършване на анализ на риска при използването на външен доставчик.

Ключови аспекти за бъдещи действия

Въпреки своите многобройни ползи цифровата ера създава и предизвикателства пред неприкосновеността на личния живот и

защитата на данните, тъй като огромни количества лична информация се събират и обработват по все по-сложни и непрозрачни начини.

Предизвикателствата пред големите данни включват тяхното регистриране, съхранение, анализ, търсене в тях, споделяне, трансфер, визуализация, правене на запитване/заявка в тях, обновяване, сигурност и видове източници за работа с тях. Важно е да се определят ясни правила и процедури за работа с големи бази данни, съдържащи лични данни на физически лица, в целия процес по тяхното обработване. Обработката на големи данни изисква нов подход, базиран на изкуствения интелект (AI) и високоскоростни мрежи.

Като интегрират защитата на данните още в проектирането на системите и процесите си и я адаптират така, че да позволява по-истинска прозрачност и потребителски контрол, отговорните администратори на данните ще могат да се възползват и от предимствата на големите масиви от данни, като същевременно осигуряват зачитане на достойнството и свободите на хората.

Цифровата грамотност и повишаването на осведомеността за цифровите права, неприкосновеността на личния живот и защитата на данните сред гражданите, включително и сред децата, както и повишаване на разбирането относно това къде и как потоците от данни се събират, са от изключително значение.

КОМИСИЯ ЗА ЗАЩИТА НА ЛИЧНИТЕ ДАННИ

бул. „Проф. Цветан Лазаров“ № 2
1592 София
Електронна поща: kzld@cpdp.bg
Интернет страница: www.cpdp.bg