

Spis treści

Przedmowa	13
Wprowadzenie	15
1. Zarządzanie danymi i proste podejście do prywatności	29
Zarządzanie danymi — co to jest?	30
Identyfikacja danych wrażliwych	32
Wskazywanie informacji umożliwiających identyfikację osoby	35
Dokumentowanie danych do wykorzystania	36
Podstawowa dokumentacja danych	36
Wyszukiwanie i dokumentowanie nieznanymi danych	41
Określanie pochodzenia danych	43
Kontrolowanie wersji danych	46
Podstawowa prywatność — pseudonimizacja	
na potrzeby ochrony prywatności w fazie projektowania	48
Podsumowanie	51
2. Anonimizacja	53
Co to jest anonimizacja?	53
Definicja prywatności różnicowej	55
Epsilon — czym jest utrata prywatności?	57
Co gwarantuje prywatność różnicowa, a czego nie?	59
Zrozumienie prywatności różnicowej	60
Prywatność różnicowa w praktyce — anonimizacja spisu powszechnego w USA	61
Prywatność różnicowa z mechanizmem Laplace'a	63
Prywatność różnicowa z rozkładem Laplace'a — podejście naiwne	65
Czułość i błąd	67
Budżety prywatności	69
Inne mechanizmy — szum gaussowski w prywatności różnicowej	71
Porównanie szumu Laplace'a i Gaussa	73
Prywatność różnicowa w świecie rzeczywistym	
— usuwanie obciążenia zaszumionych wyników	76

Jednostki czułości i prywatności	77
A co z k-anonimowością?	78
Podsumowanie	80
3. Uwzględnianie prywatności w potokach danych	81
Jak wbudować prywatność w potoki danych?	81
Zaprojektuj odpowiednie środki ochrony prywatności	82
Spotykaj się z użytkownikami tam, gdzie się znajdują	83
Implementowanie prywatności	84
Testowanie i weryfikowanie	85
Inżynieria prywatności i zarządzania danymi w potokach	85
Przykładowy przepływ pracy w udostępnianiu danych	86
Dodawanie do gromadzonych danych informacji o pochodzeniu i zgodzie	88
Wykorzystywanie bibliotek prywatności różnicowej w potokach	92
Anonimowe gromadzenie danych	96
Gromadzenie danych z prywatnością różnicową przez Apple	96
Dlaczego pierwotne zbieranie danych z prywatnością różnicową w Chrome zostało porzucone?	99
Współpraca z zespołem inżynierii danych i kierownictwem	101
Podziel się odpowiedzialnością	102
Tworzenie przepływów pracy uwzględniających dokumentowanie i prywatność	102
Prywatność jako podstawowa propozycja wartości	103
Podsumowanie	104
4. Ataki na prywatność	105
Ataki na prywatność — analiza typowych wektorów ataków	105
Atak na Netflix Prize	105
Ataki połączeniowe	108
Ataki identyfikacyjne	110
Atak na mapę Strava	111
Atak wnioskujący o członkostwo	113
Wnioskowanie o atrybutach wrażliwych	116
Inne ataki bazujące na wycieku z modelu — zapamiętywanie	117
Ataki polegające na kradzieży modeli	118
Ataki na protokoły prywatności	120
Bezpieczeństwo danych	121
Kontrola dostępu	122
Zapobieganie utracie danych	123
Dodatkowe kontrole bezpieczeństwa	123
Modelowanie zagrożeń i reagowanie na incydenty	124
Probabilistyczne podejście do ataków	125
Przeciętna osoba atakująca	125
Pomiar ryzyka i ocena zagrożeń	126

Środki zaradcze dotyczące bezpieczeństwa danych	128
Stosowanie podstawowych zabezpieczeń sieci web	128
Ochrona danych treningowych i modeli	129
Bądź na bieżąco — poznawanie nowych ataków	130
Podsumowanie	131
5. Uczenie maszynowe i nauka o danych uwzględniające prywatność	132
Wykorzystanie technik ochrony prywatności w uczeniu maszynowym	132
Techniki ochrony prywatności w typowym przepływie pracy nauki o danych lub uczenia maszynowego	133
Uczenie maszynowe chroniące prywatność w środowisku naturalnym	136
Stochastyczne zejście gradientowe z prywatnością różnicową	137
Biblioteki open source w uczeniu maszynowym chroniącym prywatność	140
Tworzenie cech z prywatnością różnicową	143
Stosowanie prostszych metod	145
Dokumentowanie uczenia maszynowego	146
Inne sposoby ochrony prywatności w uczeniu maszynowym	149
Uwzględnianie prywatności w projektach związanych z danymi i uczeniem maszynowym	152
Zrozumienie potrzeb w zakresie ochrony danych	152
Monitorowanie prywatności	153
Podsumowanie	155
6. Uczenie federacyjne i nauka o danych	156
Dane rozproszone	156
Dlaczego warto korzystać z danych rozproszonych?	157
Jak działa rozproszona analiza danych?	158
Zachowujące prywatność dane rozproszone z prywatnością różnicową	162
Uczenie federacyjne	163
Krótka historia uczenia federacyjnego	163
Dlaczego, kiedy i jak korzystać z uczenia federacyjnego	166
Projektowanie systemów federacyjnych	168
Przykładowa implementacja	169
Zagrożenia dla bezpieczeństwa	172
Przypadki użycia	173
Wdrażanie bibliotek i narzędzi federacyjnych	174
Biblioteki federacyjne typu open source	175
Flower — ujednoczony system operacyjny dla bibliotek uczenia federacyjnego	175
Przyszłość federacyjnej nauki o danych	178
Podsumowanie	178

7. Obliczenia na danych zaszyfrowanych	180
Czym są obliczenia na danych zaszyfrowanych?	180
Kiedy używać obliczeń na danych zaszyfrowanych?	181
Prywatność a tajność	183
Modelowanie zagrożeń	183
Rodzaje obliczeń na danych zaszyfrowanych	186
Bezpieczne obliczenia wielostronne	186
Szyfrowanie homomorficzne	194
Rzeczywiste zastosowania obliczeń na danych zaszyfrowanych	201
Część wspólna zbiorów prywatnych	201
Protokół Private Join and Compute	204
Bezpieczna agregacja	205
Uczenie maszynowe na danych zaszyfrowanych	206
Pierwsze kroki z PSI i Moose	207
Świat z bezpiecznym udostępnianiem danych	212
Podsumowanie	214
8. Prawna strona prywatności	215
RODO — przegląd	216
Podstawowe prawa do danych wynikające z RODO	216
Administrator danych a podmiot przetwarzający dane	218
Stosowanie zgodnych z RODO technologii zwiększających prywatność	220
Ocena skutków dla ochrony danych w RODO — zwinna i iteracyjna ocena ryzyka	223
Prawo do wyjaśnień — interpretowalność i prywatność	226
Kalifornijska ustawa o ochronie prywatności konsumentów (CCPA)	227
Stosowanie zgodnych z CCPA technologii zwiększających prywatność	228
Inne regulacje: HIPAA, LGPD, PIPL...	229
Regulacje wewnętrzne	231
Polityka prywatności i warunki korzystania z usługi	231
Umowy o przetwarzaniu danych	233
Zapoznavanie się z zasadami, wytycznymi i umowami	234
Współpraca z prawnikami	235
Przestrzeganie ustaleń umownych i prawo umów	236
Interpretacja przepisów o ochronie danych	236
Prośba o pomoc i radę	237
Wspólna praca nad definicjami i pomysłami	238
Udzielanie wskazówek technicznych	239
Zarządzanie danymi 2.0	239
Czym jest zarządzanie federacyjne?	240
Wspieranie kultury eksperymentowania	242
Działająca dokumentacja, platformy z technologią zwiększającą prywatność	243
Podsumowanie	243

9. Rozważania dotyczące prywatności i praktyczności	245
Praktyka — zarządzanie ryzykiem związanym z prywatnością i bezpieczeństwem	245
Ocena ryzyka związanego z prywatnością i zarządzanie nim	246
Uwzględnianie niepewności przy planowaniu na przyszłość	248
Technologia prywatności w praktyce — analiza przypadków użycia	251
Marketing federacyjny — prowadzenie kampanii marketingowych z wbudowaną prywatnością	251
Partnerstwa publiczno-prywatne — wymiana danych na potrzeby zdrowia publicznego	254
Zanonimizowane uczenie maszynowe — poszukiwanie zgodności z RODO w iteracyjnych ustawieniach uczenia	256
Aplikacja B2B — bez kontaktu z danymi	258
Krok po kroku — jak zintegrować i zautomatyzować prywatność w uczeniu maszynowym	259
Odkrywanie iteracyjne	260
Dokumentowanie wymagań dotyczących prywatności	261
Ocena i łączenie podejść	262
Przejsięcie na automatyzację	264
Prywatność staje się normalnością	264
Perspektywa na przyszłość — praca z bibliotekami i zespołami naukowymi	265
Współpraca z zewnętrznymi zespołami naukowymi	266
Inwestowanie w badania wewnętrzne	267
Podsumowanie	268
10. Najczęściej zadawane pytania (i odpowiedzi na nie!)	269
Obliczenia na danych zaszyfrowanych i poufne przetwarzanie danych	269
Czy obliczenia zabezpieczone są kwantowo bezpieczne?	270
Czy można używać enklaw do rozwiązywania problemów z prywatnością danych lub ich poufnością?	271
Co będzie, jeśli muszę chronić prywatność klienta lub użytkownika, który wysłał zapytanie lub żądanie do bazy danych?	271
Czy problem prywatności mogą rozwiązać clean rooms lub zdalna analiza i zdalny dostęp do danych?	272
Chcę zapewnić idealną prywatność lub idealną poufność. Czy jest to możliwe?	273
Jak ustalić, czy obliczenia na danych zaszyfrowanych są wystarczająco bezpieczne?	274
Jak zarządzać rotacją kluczy w przypadku obliczeń na danych zaszyfrowanych?	275
Czym jest piaskownica prywatności Google?	
Czy wykorzystuje obliczenia na danych zaszyfrowanych?	275
Zarządzanie danymi i mechanizmy ochrony	276
Dlaczego k-anonimowość nie jest wystarczająca?	276
Nie sądzę, by prywatność różnicowa działała w moim przypadku użycia. Co mam zrobić?	277

Czy mogę używać danych syntetycznych do rozwiązywania problemów dotyczących prywatności?	278
Jak etycznie współdzielić dane, czyli jakie są alternatywy dla sprzedaży danych?	279
Jak mogę znaleźć wszystkie prywatne informacje, które muszę chronić?	279
Po usunięciu identyfikatorów osobistych dane są bezpieczne, prawda?	280
Jak wnioskować o danych opublikowanych w przeszłości?	280
Pracuję nad pulpitem nawigacyjnym lub wizualizacją analizy biznesowej. Jak sprawić, by były przyjazne dla prywatności?	281
Kto podejmuje decyzje dotyczące inżynierii prywatności? Jak mam to wprowadzić w swojej organizacji?	282
Jakich umiejętności lub jakiego doświadczenia potrzebuję, by zostać inżynierem do spraw prywatności?	283
Dlaczego nie było mowy o (wstaw tutaj technologię lub firmę)? Jak mogę dowiedzieć się więcej? Pomocy!	284
RODO i inne przepisy o ochronie danych osobowych	284
Czy naprawdę muszę używać prywatności różnicowej do otrzymania danych niepodlegających RODO, CPRA, LGPD itp.?	285
Czy to prawda, że mogę wykorzystywać dane osobowe podlegające RODO w uzasadnionym interesie?	285
Chcę zachować zgodność ze Schrems II i transatlantyckimi przepływami danych. Jakie są możliwe rozwiązania?	286
Wybory osobiste i prywatność społecznościowa	287
Jakiego dostawcy poczty e-mail, przeglądarki i aplikacji najlepiej użyć, jeśli zależy mi na mojej prywatności?	287
Mój znajomy ma automatycznego asystenta domowego lub telefonicznego. Nie chcę, żeby mnie podsłuchiwał. Co mam zrobić?	289
Już dawno zrezygnowałem z prywatności. Nie mam nic do ukrycia. Dlaczego mam to zmienić?	290
Czy mogę po prostu sprzedać swoje dane firmom?	291
Lubię spersonalizowane reklamy. Dlaczego nie?	292
Czy (wypełnij puste miejsce) mnie podsłuchuje? Co mam z tym zrobić?	293
Podsumowanie	294

11. Idź naprzód i projektuj prywatność! 295

Kapitalizm nadzoru i nauka o danych	295
Kapitalizm GIGerów i nadzór w działaniu	296
Nadzór dla „bezpieczeństwa”	296
Luksusowy nadzór	297
Rozległe zbieranie danych i społeczeństwo	298
Uczenie maszynowe jako pranie danych	298
Dezinformacja i wprowadzanie w błąd	299

Obrona	300
Badanie, dokumentowanie, hakowanie i uczenie się	300
Kolektywizacja danych	301
Kary nakładane w związku z regulacjami	301
Wsparcie dla społeczności	302
Czempioni prywatności	303
Twoje narzędzie wielofunkcyjne do zapewniania prywatności	303
Tworzenie wiarygodnych systemów uczenia maszynowego	304
Prywatność w fazie projektowania	305
Prywatność i władza	306
Tschüss	308
Skorowidz	309