

Spis treści

Wykaz skrótów i akronimów	9
Streszczenie	11
Abstract	14
Jak korzystać z podręcznika?	17
Wstęp	21

CZĘŚĆ I. FUNDAMENTY I HISTORIA

Geneza, ewolucja i podstawy inżynierii społecznej w komunikacji cyfrowej	41
1. Historia inżynierii społecznej – od narracji wykutych w kamieniu do memów TikToka	43
2. Filary inżynierii społecznej – projektowanie zachowań zbiorowych, kontrola przepływu informacji i konstruowanie narracji	61
3. Psychologia jako narzędzie inżynierii społecznej – wykorzystanie mechanizmów poznawczych i emocjonalnych w projektowaniu przekazu	73
4. Gatekeeping 2.0 – ewolucja kontroli przepływu informacji w erze cyfrowej	87
5. Anatomia fake news – struktura, mechanizmy i rozprzestrzenianie dezinformacji	101
6. Teoria spiskowa jako produkt kulturowy – geneza, dystrybucja i funkcje w inżynierii społecznej	109
7. Komory echa i bańki filtrujące – projektowanie polaryzacji cyfrowej i granice jego skuteczności	119
8. Cambridge Analytica i profilowanie psychometryczne – studium przypadku inżynierii społecznej w kampaniach politycznych	131
9. Platform capitalism – jak modele biznesowe Big Tech i algorytmy rekomendacji projektują zachowania użytkowników	139
Podsumowanie części I	148

CZĘŚĆ II. WSPÓŁCZESNE TECHNIKI I NARZĘDZIA

Metody, technologie i praktyki inżynierii społecznej	151
10. Clickbait i sensacjonalizm	
– ekonomia uwagi w mediach cyfrowych	153
11. Viral engineering	
– mechanizmy wiralności i projektowanie zasięgów organicznych	161
12. Mikrotargeting polityczny	
– personalizacja inżynierii społecznej w mediach społecznościowych	171
13. Astroturfing i sockpuppeting	
– fałszywe tożsamości i ruchy w internecie	184
14. Trolle, farmy botów i botnety w kontekście inżynierii społecznej	
– jak państwa prowadzą wojny informacyjne	193
15. Influencer marketing jako narzędzie inżynierii społecznej	
– projektowanie autentyczności i zaufania	208
16. Dark patterns w aplikacjach	
– jak interfejsy projektują zachowania użytkowników	217
17. Nudging i architektura wyboru	
– od polityki publicznej do cyfrowego obiegu informacji	228
18. Neuromarketing i hacking uwagi	
– wykorzystanie badań mózgu w projektowaniu przekazu medialnego	239
19. Inżynieria społeczna emocji w przekazie medialnym	
– projektowanie narracji przez język, obrazy i strukturę informacji	248
20. Etyka inżynierii społecznej w dziennikarstwie	
– granica między projektowaniem przekazu a manipulacją, odpowiedzialność dziennikarska	257
Podsumowanie części II	268

CZĘŚĆ III. AI, WERYFIKACJA I OBRONA

Od automatyzacji wpływu do strategii oporu	273
21. Deepfake i synthetic media	
– inżynieria rzeczywistości w erze AI	275
22. AI i ChatGPT	
– generatywne modele językowe jako narzędzie inżynierii społecznej	286
23. Astroturfing 2.0	
– pozorowane ruchy społeczne wspomagane przez AI	296

24. Manipulacja algorytmiczna w wyszukiwarkach	
– SEO, autocomplete i sugestie jako narzędzia inżynierii społecznej	304
25. Dezinformacja wizualna	
– inżynieria przekazu przez manipulację fotografią, infografiką i montażem wideo	314
26. Memetic warfare	
– memy jako broń inżynierii społecznej w konfliktach społecznych i politycznych	321
27. Spin doctoring w erze instant news	
– inżynieria narracji w czasie rzeczywistym	330
28. Fact-checking i narzędzia weryfikacji	
– praktyka dziennikarska: InVID, TinEye, RevEye, WHOIS, Wayback Machine, analiza metadanych obrazów i wideo, wykrywanie botów	339
29. Digital literacy i programy edukacyjne	
– metody nauczania wykrywania inżynierii społecznej, budowanie odporności społecznej	351
30. Cyberhigiena, kontrpropaganda i strategie obronne dla dziennikarzy i odbiorców	
– etyczne dylematy przeciwdziałania inżynierii społecznej	360
Podsumowanie części III	369
Zakończenie	371
Aneksy	
Aneks 1. Katalog mitów	374
Aneks 2. Tabela syntetyczna technik inżynierii społecznej	378
Aneks 3. Macierz obserwowalności technik inżynierii społecznej	383
Aneks 4. Słowniczek pojęć	384
Bibliografia	390
Wykaz tabel i schematów	440
Indeks osobowy	441
Indeks rzeczowy	445
Nota o autorze	454

Streszczenie

W podręczniku przedstawiono inżynierię społeczną w komunikacji cyfrowej. Inżynieria społeczna oznacza tu świadome projektowanie i modyfikowanie zachowań zbiorowych poprzez kontrolę przepływu informacji i sterowanie środowiskiem komunikacyjnym. Komunikacja cyfrowa jest rozumiana jako komunikacja zapośredniczona przez technologie cyfrowe, zwłaszcza w tych formach, w których treść, sposób dystrybucji lub architektura środowiska komunikacyjnego służą kształtowaniu opinii, postaw i zachowań zbiorowych. Przyjęte znaczenie należy odróżnić od dwóch pozostałych znaczeń terminu: filozoficzno-politycznego, w którym inżynieria społeczna oznacza program przekształcania porządku społecznego i instytucji, oraz stosowanego w cyberbezpieczeństwie, odnoszącego się między innymi do phishingu, wyłudzenia danych i uzyskiwania dostępu do systemów.

Analiza obejmuje trzy współzależne filary inżynierii społecznej: projektowanie zachowań zbiorowych, kontrolę przepływu informacji oraz konstruowanie narracji. O zaliczeniu konkretnej praktyki do inżynierii społecznej rozstrzygają łącznie cztery cechy: świadome i systematyczne działanie, ukierunkowanie na skutek zbiorowy, projektowanie nie tylko treści, lecz także obiegu przekazu, oraz ograniczenie wspólnej obserwowalności oddziaływania. Pytanie główne brzmi: w jaki sposób narzędzia cyfrowe przekształciły warunki stosowania inżynierii społecznej?

Główne założenie podręcznika głosi, że podstawowa logika inżynierii społecznej wykazuje ciągłość historyczną. Mechanizmy wpływu znane z propagandy minionego wieku pozostają rozpoznawalne w komunikacji cyfrowej, ponieważ opierają się na względnie trwałych właściwościach ludzkiego poznania. Przekształceniu uległo natomiast środowisko jej stosowania: skala i szybkość dystrybucji, koszt krańcowy dotarcia do kolejnego odbiorcy, szczegółowość segmentacji, możliwości personalizacji, automatycznego testowania i optymalizacji przekazu oraz stopień jego publicznej obserwowalności. W rezultacie oddziaływanie przesunęło się od masy ku coraz węższym segmentom, a potencjalnie ku pojedynczemu odbiorcy, zachowując jednocześnie masowy zasięg.

Autor nie zakłada, że przekaz cyfrowy skuteczniej zmienia postawy niż propaganda prasowa, radiowa lub telewizyjna. Rozróżnienie warunków oddziaływania i jego skutków organizuje cały wywód. Sam fakt zastosowania określonej techniki nie stanowi jeszcze dowodu jej skuteczności. Udokumentowanie, że technikę zastosowano, oraz wykazanie, że rzeczywiście wpłynęła ona na opinie, postawy lub zachowania odbiorców, należą do dwóch odmiennych porządków

dowodowych. Brak potwierdzenia skuteczności nie jest przy tym dowodem nieskuteczności. Również reakcje rejestrowane przez platformy, takie jak kliknięcia, udostępnienia lub czas oglądania, nie stanowią automatycznie dowodu zmiany opinii ani zachowania poza platformą.

Podstawą opracowania jest krytyczna, problemowa analiza literatury i źródeł, obejmująca publikacje naukowe, akty prawne, dokumenty instytucjonalne, raporty organizacji wyspecjalizowanych oraz udokumentowane studia przypadków. Zasadą prowadzenia analizy jest demitologizacja, rozumiana jako krytyczna weryfikacja obiegowych twierdzeń o cyfrowej manipulacji, przede wszystkim o jej sile i skuteczności. Pozwala ona oddzielić ustalenia znajdujące oparcie w badaniach od sensacyjnych narracji, przekonań obiegowych i deklaracji podmiotów zainteresowanych wyolbrzymianiem skuteczności własnych narzędzi. Demitologizacja jest prowadzona symetrycznie. Podręcznik odrzuca zarówno technopanikę, przedstawiającą cyfrową manipulację jako niemal wszechmocną, jak i bagatelizację, według której technologie cyfrowe nie przyniosły żadnej istotnej zmiany. Nie zmienił się człowiek. Zmieniły się warunki, w których można do niego docierać, obserwować jego reakcje, testować warianty przekazu i różnicować kierowane do niego komunikaty.

Podręcznik pełni trzy zasadnicze funkcje. Po pierwsze, porządkuje rozproszone pojęcia, techniki i mechanizmy inżynierii społecznej. Po drugie, krytycznie oddziela ustalenia potwierdzone od mitów, technopaniki i twierdzeń obiegowych. Po trzecie, dostarcza studentom, nauczycielom, dziennikarzom, twórcom i odbiorcom przekazu cyfrowego, osobom prowadzącym komunikację publiczną oraz autorom kampanii politycznych i społecznych narzędzi rozumienia i rozpoznawania działań służących projektowaniu zachowań zbiorowych.

Autorski rdzeń podręcznika stanowi model łączący cztery procesy opisywane w piśmiennictwie najczęściej osobno: platformizację gatekeepingu, personalizację, automatyzację i skalę. Zostają one ujęte jako nakładające się aspekty jednego przekształcenia – ograniczenia wspólnej obserwowalności przekazu. Gatekeeping nie zanikł, lecz został przekształcony i częściowo przejęty przez platformy, systemy rekomendacyjne, moderatorów, użytkowników oraz reklamodawców. Platformizacja utrudnia rozpoznanie kryteriów selekcji treści. Personalizacja osłabia wspólność komunikatu, ponieważ poszczególni odbiorcy mogą otrzymywać jego odmienne warianty. Automatyzacja rozprasza sprawstwo między organizatora kampanii, projektanta systemu, platformę i model algorytmiczny. Skala utrudnia natomiast odtworzenie całości operacji, gdyż liczba wariantów, emisji i interakcji może przekraczać możliwości pojedynczego obserwatora.

Obserwowalność pełni funkcję kategorii opisowej. Odnosi się do możliwości zobaczenia przekazu, porównania jego wariantów, ustalenia kryteriów dystrybucji oraz przypisania odpowiedzialności za jego emisję. Jawność ma natomiast charakter normatywny i oznacza wymaganie, aby zasadniczy cel, źródło i sposób

oddziaływania nie były ukrywane przed odbiorcą. Rozpatrywana łącznie z prawdziwością przekazu, poszanowaniem autonomii odbiorcy i odpowiedzialnością nadawcy, współwyznacza granicę między dziennikarstwem a manipulacją. Przekaz ma charakter manipulacyjny zwłaszcza wtedy, gdy jego działanie zależy od nierozpoznania przez odbiorcę intencji nadawcy lub zastosowanej techniki wpływu.

Materiał został uporządkowany w trzydziestu problemowych tematach podzielonych na trzy części. Wywód zachowuje formę pytań i odpowiedzi, krótkich definicji, przykładów, kontrastów oraz praktycznych wskazówek. Każdy temat zamyka stały blok dydaktyczny, obejmujący ćwiczenie, jednozdaniowy rdzeń tematu, pytania kontrolne i wskazówki bibliograficzne. Część pierwsza obejmuje fundamenty i historię inżynierii społecznej, wskazując ciągłość jej mechanizmów psychologicznych oraz przekształcenie warunków ich stosowania. Część druga analizuje współczesne techniki i narzędzia oraz krytycznie weryfikuje przypisywane im skutki. Część trzecia dotyczy sztucznej inteligencji jako czynnika automatyzującego znane techniki wpływu oraz narzędzi weryfikacji i strategii oporu dostępnych dziennikarzowi i odbiorcy.

Generatywna sztuczna inteligencja zostaje przedstawiona jako kolejny etap tego samego procesu. Nie tworzy całkowicie nowych podstaw ludzkiej podatności na wpływ, lecz zwiększa skalę, szybkość, automatyzację i możliwości personalizacji przekazu. Pozwala generować liczne warianty komunikatu, dostosowywać je do odbiorcy, symulować nadawcę oraz prowadzić interakcję w czasie rzeczywistym. W tych obszarach zmiana ilościowa może zostać doprowadzona tak daleko, że zaczyna ocierać się o zmianę jakościową.

Z przeprowadzonej diagnozy wynika podstawowy kierunek obrony. Nie wystarczy prostować każdy fałsz osobno ani koncentrować się na kolejnych narzędziach, które szybko mogą zostać zastąpione nowymi. Odpowiedź wymaga odbudowy warunków, w których przekaz można zobaczyć, porównać, przypisać nadawcy i poddać publicznej ocenie. Obrona przed współczesną inżynierią społeczną jest więc w znacznej mierze obroną obserwowalności komunikacji. Przedstawiony w podręczniku stan badań, technologii i regulacji odpowiada lipcowi 2026 roku.

Słowa kluczowe: inżynieria społeczna, komunikacja cyfrowa, obserwowalność przekazu, platformizacja, personalizacja, mikrotargeting, dezinformacja, dziennikarstwo, manipulacja, jawność, sztuczna inteligencja.

From the Masses Towards the Individual Social Engineering in Digital Communication

Abstract

This textbook examines social engineering in digital communication. Social engineering is understood here as the deliberate design and modification of collective behaviour through control over information flows and the steering of the communication environment. Digital communication is understood as communication mediated by digital technologies, particularly in those forms in which the content, mode of distribution or architecture of the communication environment serves to shape collective opinions, attitudes and behaviour. This meaning must be distinguished from the two other senses of the term: the philosophical and political one, in which social engineering denotes a programme of transforming the social and institutional order, and the one used in cybersecurity, which includes phishing, obtaining data by deception and gaining unauthorised access to systems.

The analysis encompasses three interdependent pillars of social engineering: the design of collective behaviour, control over information flows and the construction of narratives. A given practice qualifies as social engineering only if four features are jointly present: deliberate and systematic action, an orientation towards collective effects, the design not only of content but also of message distribution, and limited shared observability of the influence exercised. The principal question guiding the textbook is: how have digital tools transformed the conditions under which social engineering is applied?

The textbook's central premise is that the underlying logic of social engineering exhibits historical continuity. Mechanisms of influence known from the propaganda of the past century remain recognisable in digital communication because they rest on relatively stable properties of human cognition. What has changed is the environment in which it is practised: the scale and speed of distribution, the marginal cost of reaching an additional recipient, the granularity of segmentation, the scope for personalisation, automated testing and message optimisation, and the degree to which messages are publicly observable. As a result, influence shifts from the mass towards increasingly narrow segments and, potentially, towards the individual recipient, while retaining mass reach.

The textbook does not assume that digital messages change attitudes more effectively than press, radio or television propaganda. The distinction between the conditions of influence and its effects structures the entire argument. The mere use of

a particular technique does not constitute evidence of its effectiveness. Documenting that a technique was used and demonstrating that it actually influenced recipients' opinions, attitudes or behaviour belong to two distinct evidentiary domains. A lack of evidence confirming effectiveness is not evidence of ineffectiveness. Nor do platform-recorded responses, such as clicks, shares or viewing time, automatically constitute evidence of changes in opinion or behaviour beyond the platform.

The textbook draws on a critical, problem-oriented analysis of literature and sources, including scholarly publications, legal acts, institutional documents, reports by specialised organisations and documented case studies. The analysis is guided by demythologisation, understood as the critical examination of commonly held claims about digital manipulation, and above all about its power and effectiveness. This makes it possible to distinguish findings supported by research from sensational narratives, commonplace assumptions and claims made by actors with an interest in exaggerating the effectiveness of their own tools. Demythologisation is conducted symmetrically. The textbook rejects both technopanic, which portrays digital manipulation as virtually omnipotent, and trivialisation, according to which digital technologies have introduced no significant change. Human beings have not changed. What has changed are the conditions under which people can be reached, their reactions observed, message variants tested and the messages directed at them differentiated.

The textbook performs three principal functions. First, it organises the dispersed concepts, techniques and mechanisms of social engineering. Second, it critically distinguishes established findings from myths, technopanic and commonplace claims. Third, it provides students, teachers, journalists, creators and recipients of digital content, those engaged in public communication and the authors of political and social campaigns with tools for understanding and recognising practices designed to shape collective behaviour.

At the core of the textbook is an original model combining four processes that are most often discussed separately in the literature: the platformisation of gatekeeping, personalisation, automation and scale. They are conceptualised as overlapping aspects of a single transformation: the limitation of the shared observability of messages. Gatekeeping has not disappeared; it has been transformed and partly taken over by platforms, recommendation systems, moderators, users and advertisers. Platformisation makes the criteria governing content selection more difficult to identify. Personalisation weakens the commonality of the message because different recipients may receive different versions of it. Automation disperses agency among the campaign organiser, the system designer, the platform and the algorithmic model. Scale, in turn, makes it more difficult to reconstruct the operation as a whole, since the number of message variants, impressions and interactions may exceed the capacities of any single observer.

Observability functions as a descriptive category. It refers to the possibility of seeing a message, comparing its variants, identifying the criteria governing its distribution and attributing responsibility for its dissemination. Transparency, by contrast, is a normative requirement: the essential purpose, source and mode of influence must not be concealed from the recipient. Considered together with the truthfulness of the message, respect for the recipient's autonomy and the accountability of the sender, this requirement helps define the boundary between journalism and manipulation. A message is manipulative especially when its effect depends on the recipient's failure to recognise the sender's intention or the influence technique being used.

The material is organised into thirty problem-oriented topics divided into three parts. The argument retains the form of questions and answers, brief definitions, examples, contrasts and practical guidance. Each topic closes with a didactic block comprising an exercise, a one-sentence takeaway, self-check questions and suggestions for further reading. The first part examines the foundations and history of social engineering, demonstrating the continuity of its psychological mechanisms and the transformation of the conditions in which they are applied. The second analyses contemporary techniques and tools and critically assesses the effects attributed to them. The third addresses artificial intelligence as a factor automating known techniques of influence, together with the verification tools and strategies of resistance available to journalists and audiences.

Generative artificial intelligence is presented as a further stage in the same process. It does not create entirely new foundations of human susceptibility to influence, but increases the scale, speed and automation of message production and expands the scope for personalisation. It can generate numerous variants of a message, tailor them to particular recipients, simulate a sender and sustain interaction in real time. In these areas, quantitative change may be taken so far that it begins to verge on qualitative change.

This diagnosis points to a fundamental direction for defence. It is not enough to correct each falsehood individually or to focus on successive tools that may soon be replaced by new ones. An effective response requires restoring the conditions under which messages can be seen, compared, attributed to a sender and subjected to public scrutiny. Defence against contemporary social engineering is therefore, to a significant extent, a defence of the observability of communication. The textbook reflects the state of research, technology and regulation as of July 2026.

Keywords: social engineering, digital communication, message observability, platformisation, personalisation, microtargeting, disinformation, journalism, manipulation, transparency, artificial intelligence.