

Witold Wrotek

Eleven*Labs*

Zastosowanie w biznesie i twórczości

Wszelkie prawa zastrzeżone. Nieautoryzowane rozpowszechnianie całości lub fragmentu niniejszej publikacji w jakiegokolwiek postaci jest zabronione. Wykonywanie kopii metodą kserograficzną, fotograficzną, a także kopiowanie książki na nośniku filmowym, magnetycznym lub innym powoduje naruszenie praw autorskich niniejszej publikacji.

Wszystkie znaki występujące w tekście są zastrzeżonymi znakami firmowymi bądź towarowymi ich właścicieli.

Autor oraz wydawca dołożyli wszelkich starań, by zawarte w tej książce informacje były kompletne i rzetelne. Nie biorą jednak żadnej odpowiedzialności ani za ich wykorzystanie, ani za związane z tym ewentualne naruszenie praw patentowych lub autorskich. Autor oraz wydawca nie ponoszą również żadnej odpowiedzialności za ewentualne szkody wynikłe z wykorzystania informacji zawartych w książce.

Redaktor prowadzący: Małgorzata Kulik

Projekt okładki: Studio Gravite/Olsztyn
Obarek, Pokoński, Pazdrijowski, Zaprucki

Grafika na okładce została wykorzystana za zgodą AdobeStock.com.

Helion S.A.
ul. Kościuszki 1c, 44-100 Gliwice
tel. 32 230 98 63
e-mail: helion@helion.pl
WWW: helion.pl (księgarnia internetowa, katalog książek)

Drogi Czytelniku!
Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres
helion.pl/user/opinie/elevla
Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

ISBN: 978-83-289-3488-7

Copyright © Helion S.A. 2026

Printed in Poland.

- [Kup książkę](#)
- [Poleć książkę](#)
- [Oceń książkę](#)

- [Księgarnia internetowa](#)
- [Lubię to! » Nasza społeczność](#)

SPIIS TREŚCI

1 ELEVENLABS — SZYBKIE PRZEWODNIKI DLA POCZĄTKUJĄCYCH	13
Dlaczego Voice AI i dlaczego ElevenLabs	13
Założ konto	13
Interfejs w skrócie	14
Pierwsze nagranie lektorskie w 15 minut	14
Formaty eksportu i ich zastosowanie	15
Mała dawka inżynierii promptów	16
Bezpieczeństwo, etyka i warunki korzystania (wersja skrócona — 60 sekund)	16
Najczęściej zadawane pytania dotyczące rozwiązywania problemów	17
2 JAK W ZALEDWIE 15 MINUT STWORZYĆ SWÓJ PIERWSZY GŁOSOWY KOMENTARZ Z WYKORZYSTANIEM TECHNOLOGII SYNTEZY MOWY	18
Wyzwanie, któremu sprostasz w zaledwie 15 minut	18
Warsztat doskonalący: od tweeta do minutowego podcastu	21
Najczęściej napotykanne trudności oraz sprawdzone metody ich rozwiązania	22
Najważniejsze wnioski	22
Bibliografia	23
3 PLANY BEZPŁATNE A PŁATNE: JAK WYBRAĆ WŁAŚCIWĄ SUBSKRYPCJĘ WSPIERAJĄCĄ ROZWÓJ	24
Co tak naprawdę kupujesz: jednostki rozliczeniowe, minuty, prawa i funkcje	24
Pięć profili twórców — wybierz ścieżkę odpowiednią dla siebie	25
Trójetapowe drzewo decyzyjne (60 sekund)	26
Co się zmienia po przejściu na wyższy poziom, czyli „odczucie” każdego progu	26
Prawa, zasady i zgodność (przeczytaj przed publikacją)	27
Szacowanie wolumenu: minuty, generowanie i równoczesność	28
Planowanie budżetu i momenty przejścia na wyższy plan (praktyczne progi)	29
Realia platformy, które wpływają na wybór planu	30

Praktyczny wybór planu: jednokartkowa macierz porównawcza	30
Lista kontrolna zgodności (do wklejenia w procedurę operacyjną)	31
Sedno sprawy	31
Bibliografia	32
4 JAK BEZPIECZNIE I ZGODNIE Z PRAWEM SKLONOWAĆ GŁOS — SWÓJ LUB CUDZY	33
Dlaczego to ma dziś szczególne znaczenie	33
Trzy filary bezpiecznego klonowania głosu — tzw. trzy C	34
Co tak naprawdę interesuje prawo (w języku zrozumiałym dla każdego)	34
Klonowanie głosu: mapa ryzyka — od bezpiecznego do ryzykownego	36
Praktyczny przewodnik od A do Z —	
jak bezpiecznie, przejrzysto i zgodnie z prawem klonować głos	36
Jednostronicowa zgoda na wykorzystanie głosu i jego klonowanie	
(wersja skrócona)	38
Przypadki szczególne, które napotkasz w praktyce	40
Dobre praktyki nagraniowe — by Twój sklonowany głos	
brzmiał świetnie i bezpiecznie	41
Oznaczanie i znakowanie, które naprawdę mają sens	41
Lista kontrolna „Bez żalu” — warto wydrukować	42
Czego nigdy nie wolno robić	42
Dwie przestrogi z życia wzięte	43
Szybkie FAQ, czyli najczęstsze pytania i odpowiedzi	43
Słowo na zakończenie	44
Bibliografia	44
5 PRECYZYJNE WSKAZÓWKI NAGRANIOWE	
DLA MODELI GŁOSU WYSOKIEJ WIERNOŚCI	47
Co naprawdę oznacza „wysoka wierność” w modelach głosu	47
Pomieszczenie: najpierw zadbaj o przestrzeń, potem o sprzęt	48
Wybór mikrofonu, charakterystyka kierunkowa i ustawienie	48
Zawsze czysty sygnał wejściowy: wzmocnienie, rozdzielczość bitowa	
i częstotliwość próbkowania	50
Dwuminutowy rytuał przygotowawczy — powtarzaj przed każdą sesją	50
Jak przygotować tekst do nagrań szkoleniowych — sztuka skryptu	51
Technika „złotego ujęcia” — szybka gwarancja jakości	51
Zarządzanie plikami na dużą skalę — porządek, który procentuje	52
Szybka lista kontrolna jakości nagrania (5 minut uwagi)	52
Rozwiązywanie problemów: przyczyna → rozwiązanie	53
Co sprzyja modelowi głosu, a czego należy unikać	54
Zaawansowane dodatki — gdy zależy Ci na ostatnich 5% jakości	54
Protokół powtarzalności — do wklejenia w procedurę operacyjną (SOP)	55
Myśli na zakończenie	56
Bibliografia	56

6 KŁONOWANIE GŁOSU W WIELU JĘZYKACH I Z WIELOMA AKCENTAMI — GLOBALNY ZASIĘG JEDNYM KLIKNIĘCIEM	58
Co rzeczywiście przenosi się między językami, a co nie	58
Sprawdzenie możliwości: czy Twój model „zna” dany język?	59
Pięcioletniowy schemat klonowania głosu wielojęzycznego i akcentowego	59
Dialekt, akcent i rejestr — czym się różnią i jak je kontrolować	60
Kontrola wymowy — lekkie narzędzia, które naprawdę działają	61
Kontrola jakości z udziałem rodzimych użytkowników języka — szybko i rzetelnie	62
Przełączanie kodów językowych i nazwy marek	62
Prawo i przejrzystość ponad granicami	63
Dlaczego mowa międzyjęzykowa jest trudna, choć coraz łatwiejsza	63
Jak radzić sobie z pułapkami językowymi i akcentowymi	64
Wzorce dostarczania treści, które łatwo skalować	64
Jednostronicowa procedura operacyjna — do skopiowania i wdrożenia	65
Bibliografia	66
7 SZTUKA KONSTRUOWANIA PROMPTÓW DLA NATURALNIE BRZMIĄCEJ NARRACJI AI	68
Co tak naprawdę robi prompt, czyli kulisy działania	68
Narracyjna struktura promptu, czyli pięć linijek, które robią różnicę	69
Najpierw pisz dla ucha — drobne poprawki, które robią wielką różnicę	69
SSML — trzy znaczniki, które warto znać na co dzień	70
Kontrola wymowy — nazwy własne, marki i zapożyczenia	71
Dwanaście uniwersalnych wzorców promptów — kopiuj, dostosuj, publikuj	71
Kontrola prozodii poza znacznikami — gdy zależy Ci na subtelności	74
Liczby, daty i skróty — niech zabrzmiały po ludzku	74
Tempo dla zrozumienia, nie dla pośpiechu	75
Siedmiominutowa kontrola jakości — powtarzalna i zespołowa	76
Diagnoza i korekta — od promptu, przez objaw, do rozwiązania	76
Przepis w pigułce: jak z tekstu płaskiego stworzyć żywą narrację — w trzech krokach	77
Stwórz własną dokumentację stylu, by przyszłe nagrania brzmiały tak dobrze, jak obecne	78
Bibliografia	79
8 KOMPLETNY PRZEWODNIK PO SSML — CZYLI TON, TEMPO I EMOCJA W NARRACJI GŁOSOWEJ	80
SSML w minutę — czym jest i dlaczego ma znaczenie	80
Szablon minimalistyczny, czyli jak zacząć z SSML	81
Tempo, wysokość i głośność, czyli magia <prosody>	81
Mikrostrategie, które naprawdę działają	82

Naturalne pauzy z użyciem znacznika <break>	83
Podkreślenie bez krzyku — czyli jak mówić wyraziście, lecz z klasą	83
Spraw, by silnik czytał tak, jak robią to ludzie — czyli magia znacznika <say-as>	84
Nazwy własne, nazwy marek i trudne słowa — czyli jak mówić poprawnie dzięki <phoneme> i <sub>	85
Zmiana języka i głosu — czyli jak mówić wielojęzycznie z klasą	86
Struktura i synchronizacja — czyli jak mówić z oddechem i rytmem	87
Emocje i styl — co jest standardem, a co dodatkiem od dostawcy	87
Szybka mapa zgodności między dostawcami — warto mieć w notatniku	89
Przepisy na „mikroznaczniki” — czyli jak nadać mowie styl, tempo i emocje	89
Testowanie i kontrola jakości w pięć minut — czyli jak mówić poprawnie, nie tracąc czasu	91
Typowe pułapki → szybkie poprawki	92
Systemy stylu i „emocji” — głębsza warstwa wyrazu	92
Gdy SSML to za mało	93
Zachowaj przenośność — krótka instrukcja dobrej praktyki	93
Bibliografia	94

9 SERYJNE GENEROWANIE DIALOGÓW

Z UŻYCIEM API ELEVENLABS I JĘZYKA PYTHON	96
Co zbudujesz i dlaczego to się skaluje	96
Co otrzymasz po zakończeniu pracy	96
Wymagania wstępne i konfiguracja środowiska	97
Format pliku z dialogami — prosty, lecz solidny	98
Dialogowy „Hello World” w 20 liniach (jedna linia → jeden plik)	98
Zachowanie płynności prozodii przy dzieleniu tekstu na fragmenty (pola kontekstowe)	99
Produkcyjna linia przetwarzania dialogów — asynchronicznie i z klasą	100
Pozyskiwanie znaczników czasowych do napisów i precyzyjnego montażu scen	102
Łączenie klipów w jedną scenę — czysty Python	103
Dobór odpowiedniego modelu, formatu i języka	104
Mapowanie głosów postaci i wymowa — na dużą skalę	105
Łączenie zapytań i poprawki — zachowanie spójności	105
Praca w czasie rzeczywistym i strumieniowanie — gdy tryb wsadowy nie wystarcza	105
A co z interfejsem Text-to-Dialogue?	106
Porady operacyjne — by Twoje procesy nigdy się nie zatrzymały	106
Lista kontrolna jakości — szybko i rzetelnie	106
Krok dalej — miksowanie, napisy, automatyzacja	107
Bibliografia	107

10	YOUTUBE I TIKTOK — PRACA Z LEKTOREM AI NA KANAŁACH „BEZ TWARZY”	109
	Czym powinien być kanał „bez twarzy” i jak powinien brzmieć	109
	Kompletny przebieg pracy — spojrzenie z lotu ptaka	110
	Od scenariusza do głosu AI w kilka minut — przepis, który warto powtarzać	110
	Podstawy dźwięku — by widzowie chcieli słuchać	111
	Obraz — formaty, które nie sprawią Ci kłopotu później	112
	Lista kontrolna przed publikacją — by nie trzeba było eksportować ponownie	112
	Zasady, które chronią monetyzację	113
	Praktyczny dzień produkcyjny — YouTube (forma dłuższa)	115
	Praktyczna godzina produkcyjna — Shorts/TikTok	116
	Gotowe szablony, które możesz wkleić już dziś	116
	Napisy — dostępność, zasięg i czas oglądania	117
	Retencja — popraw pierwsze 30 sekund	117
	Muzyka bez pułapek	117
	Zasady kluczowe — na jednej stronie	118
	Rozwiązywanie problemów — czyli jak nie dać się zaskoczyć	118
	Skalowanie za pomocą systemów	118
	Procedura produkcyjna dla dwóch platform — gotowa do skopiowania	119
	Bibliografia	120
11	PODCAST BEZ MIKROFONU — OD SCENARIUSZA DO EMISJI W GODZINĘ	122
	Mapa drogowa w godzinę — od pomysłu do podcastu	122
	Związły i powtarzalny szablon scenariusza	124
	Generowanie naturalnego głosu AI — bez mikrofonu	124
	Montaż: muzyka, efekty i płynność — nadal bez mikrofonu	125
	Mastering zgodny ze specyfikacją platform — by Twój podcast brzmiał „jak należy”	126
	Okładka i notatki do odcinka — zgodne z wymaganiami katalogów	127
	Eksport, hosting i dystrybucja — czyli co kliknąć, by podcast trafił do świata	128
	Kontrola jakości — trzyminutowa lista przed premierą	130
	Szybkie rozwiązania typowych potknięć	130
	Uwagi etyczne i prawne — warte minuty skupienia	132
	Twój pierwszy odcinek bez mikrofonu — przykład krok po kroku	132
	Porady dla naturalnej, „ludzkiej” narracji AI	133
	Monetyzacja (opcjonalna już od pierwszego dnia)	133
	Plan sezonu „bez mikrofonu”	134
	Jak zachować zgodność podczas rozwoju podcastu	134
	Bibliografia	134

12 PRODUKCJA AUDIOBOOKA DIY DLA AUTORÓW KINDLE I PLATFORMY ACX	137
Mapa audiobooka w 90 sekund	137
Wybierz model dystrybucji (by nie powtarzać pracy później)	139
Etap przedprodukcyjny (dla ACX i innych platform)	140
Nagrywanie audiobooka (narracja ludzka)	141
Edycja i korekta (czyli co naprawdę wychwytuje kontrola jakości ACX)	142
Mastering dla ACX — liczby, które warto znać, osiągnąć i... zapomnieć	142
Pakowanie, które przechodzi za pierwszym razem	144
Publikacja na ACX i wybór modelu tantiem	144
Równoległa edycja z narracją AI — gdzie i jak ją opublikować	145
Dobór lektora i współpraca z nim (jeśli nie chcesz czytać samodzielnie)	146
Jednodniowy plan produkcji (narracja własna)	147
Lista kontrolna jakości (do skopiowania i wykorzystania)	148
Typowe błędy i szybkie sposoby ich naprawy	149
Realistyczne spojrzenie na kwestie prawne i regulacyjne	149
Matematyka premiery — czyli jak zaplanować zwrot z inwestycji (ROI)	150
Prosty i powtarzalny schemat działania (ścieżka ACX)	151
Prosty i powtarzalny schemat działania (ścieżka przyjazna AI)	152
Ostatni impuls	152
Bibliografia	153
13 MODYFIKACJE GIER I ŚWIATY WIRTUALNE — JAK DODAĆ REALISTYCZNE GŁOSY POSTACI NIEZALEŻNYCH (NPC)	155
Wybierz strategię głosową: offline czy real-time?	155
Projektowanie wiarygodności — co warto udźwiękować, kiedy i w jaki sposób	157
Praktyczna ścieżka offline — „wypiecz raz, używaj wszędzie”	158
Praktyczna ścieżka w czasie rzeczywistym — dynamiczna mowa, minimalne obciążenie pamięci	160
Unity — gotowe wzorce, które sprawdzają się w produkcji	161
Unreal Engine — solidne podejścia do mowy strumieniowanej i „wypiekanej”	162
Bethesda Creation Kit (Skyrim, Fallout, Starfield): głos, synchronizacja ust (LIP) i pakowanie (FUZ)	163
Serwery FiveM (GTA V): odtwarzanie mowy NPC w świecie gry	164
Mody Minecrafta (Forge/Fabric): zdarzenia dźwiękowe przez paczki zasobów	164
Second Life i VRChat — światy społeczne z zasadami	165
Napisy, podpisy i dostępność — czyli jak mówić do wszystkich	166
Synchronizacja ust — od „wystarczająco dobrze” po „efekt wow”	166
Miksowanie, wydajność i budżet pamięci	167
Kontrola jakości, która naprawdę wychwytuje błędy	168

Polityka, zgoda i rzeczywistość — nie pomijaj tego	168
Gotowe schematy, które możesz wykorzystać	169
Szybkie rozwiązania typowych problemów	170
Dokąd zmierza branża — kontekst dla Twojej mapy drogowej	170
Ostatnia uwaga	171
Bibliografia	171
14 FIVERR, UPWORK I NIE TYLKO — JAK SPRZEDAWAĆ USŁUGI LEKTORSKIE	174
Gdzie sprzedawać (i dlaczego)	174
Oferty, które przekonują: sprzedawaj efekty, nie minuty	175
Przejrzysta i etyczna sztuczna inteligencja:	
ujawnienie, które buduje zaufanie (i chroni Cię prawnie)	176
Jak szybko wybrać swoją niszę	177
Twój showreel — jak pokazać wartość w 60 sekund	177
Ceny, które mają sens — dla Ciebie i dla klienta	178
Prawa użytkownika i licencjonowanie — jasno i przejrzysto	179
SEO dla ofert na Fiverrze i sztuka pisania propozycji na Upworku	180
Proces produkcji, który rośnie wraz z Tobą i pozostaje spójny	181
Polityka poprawek —	
jak uniknąć niekontrolowanego rozszerzania zakresu	182
Opinie i dowody społeczne — zgodnie z prawem	182
Prawa, zgody i treści wrażliwe — jak postępować odpowiedzialnie	183
Sprzedaż poza główną ofertą — jak zwiększyć wartość zamówienia	183
Jak ograniczyć ryzyko związane z platformami (opłaty, oferty, anulacje)	183
Plan startowy na 30 dni — lista działań	184
Szablony, które warto wykorzystać	185
Negocjacje i komunikacja — nawyki skutecznych wykonawców	186
Wskaźniki, które naprawdę się liczą (i jak je poprawić)	186
Czerwone flagi — poznaj je teraz, by uniknąć problemów później	187
Twoja przewaga: niezawodność + wielojęzyczność + przejrzystość	187
Bibliografia	188
15 STRUMIENIE DOCHODU PASYWNEGO —	
AUDIOBOOKI, KURSY I PĘTLE REKLAMOWE	191
Koło zamachowe: jeden tekst → trzy aktywa	191
Aktywum nr 1 — audiobooki (ACX + platformy przyjazne AI)	192
Aktywum nr 2 — kursy online (platformy edukacyjne vs. własna szkoła)	194
Aktywum nr 3 — pętłe przychodów reklamowych	
(YouTube, Shorts, TikTok)	195
Mapy finansowe — prosta matematyka, której możesz zaufać	197
Wykorzystanie treści — tygodniowy sprint produkcyjny	197
Architektura skalowania — by Twoje pętłe nie pękały	198

Lejki startowe — kopiuj i dostosuj	199
Zgodność i zaufanie — kwestie niepodlegające negocjacji	199
Plan budowy w 30 dni	200
Najczęściej zadawane pytania — te finansowe	201
Myśl końcowa	202
Bibliografia	202
16 PRAWO AUTORSKIE, ZGODA I REGULACJE DOTYCZĄCE DEEPFAKE'ÓW — WYJAŚNIONE PRZYSTĘPNIE	205
Dlaczego to dziś tak istotne	205
Podstawy prawa autorskiego — na jednej stronie (i dlaczego nadal mają zastosowanie do głosów AI)	206
Głos, wizerunek i zgoda — czyli prawo do komercyjnego wykorzystania podobizny	207
Deepfaki i przepisy dotyczące mediów syntetycznych — z którymi faktycznie się zetkniesz	208
Zasady platform, z którymi się zetkniesz (YouTube, TikTok, Meta, Spotify, ElevenLabs)	209
Znakowanie i pochodzenie treści — jak udowodnić, że nagranie jest syntetyczne (lub autentyczne)	210
Praktyczne listy zgodności — do wklejenia w swój workflow	210
Scenariusze, dane treningowe i pytanie: „Kto jest właścicielem efektów pracy AI?”	211
Przykłady zastosowania — głos syntetyczny w praktyce	212
Proste zgody i pliki projektowe — krótkie, zrozumiałe komponenty	213
Pytania najczęściej zadawane przez twórców	214
Mapa „czerwonych flag” dla twórców — kiedy warto się zatrzymać i skonsultować z prawnikiem	215
Jednostronicowy szablon zgodności — do uzupełnienia	215
Sposób myślenia, który chroni i wspiera twórczość	216
Bibliografia	216
17 UMOWY I FORMULARZE ZGODY — WZORCE DOBREJ PRAKTYKI	220
Dlaczego umowy mają szczególne znaczenie w przypadku głosów AI	220
Zestaw umów dla projektów głosowych z AI — co podpisać i kiedy	221
Złote klauzule — co oznaczają i jak je sformułować	222
Zgoda lektora i aneks AI (wzór)	225
Aneks dla osoby małoletniej (wzór)	227
Umowa ramowa o świadczenie usług (wersja skrócona — wzór)	227
Aneks dotyczący przetwarzania danych (klauzule zasadnicze)	228
Umowa o poufności (NDA, dwustronna) — wersja uproszczona	229
Zgoda na wykorzystanie opinii/recenzji — zgodna z zasadami FTC	229
Potwierdzenie licencji na muzykę, efekty dźwiękowe i zasoby	230

Lista kontrolna podpisu elektronicznego (zgodna z ESIGN/eIDAS)	230
Uwagi międzynarodowe (USA/UE/UK)	230
Pochodzenie i ujawnienie — opcjonalne, lecz rozsądne	231
Taktyka negocjacyjna — podręcznik w pigułce	231
Jednostronicowy „Pakiet projektowy” — umieść w każdym folderze zlecenia	232
Często zadawane pytania — „A co, jeśli...?”	232
Zestaw gotowych miniszablonów (do skopiowania, dostosowania i użycia)	233
Lista kontrolna jakości końcowej (wydrukuj i dołącz do projektu)	234
Bibliografia	235
18 OCHRONA GŁOSU MARKI PRZED NADUŻYCIAMI	239
Zagrożenie w języku prostym	239
Twój system obronny: P•D•R•E	240
ZAPOBIEGAJ — zbuduj wielowarstwową tarczę ochronną	241
WYKRYWAJ — wczesne ostrzeganie bez zespołu SOC	243
REAGUJ — plan działania na 72 godziny	243
EGZEKWUJ — dobierz właściwe narzędzie prawne	245
Przypadki szczególne	245
Szablony polityk i procedur (do kopiowania, dostosowania i wdrożenia)	246
Miarodajne wskaźniki	247
Szybkie pytania i odpowiedzi	248
Bibliografia	248
19 SZTUCZKI Z DOPASOWYWANIEM SZUMU, TRANSFEREM AKCENTU I WYGŁADZANIEM DŹWIĘKU	251
Dlaczego nawet „znakomite” głosy AI brzmią sztucznie i jak temu zaradzić	251
Dopasowywanie szumu: jak osadzić głos w przestrzeni	252
Transfer akcentu, który nie burzy wiarygodności	254
Wygładzanie dźwięku: sprawdzony łańcuch mistrzowski	256
Łączenie wszystkiego w całość: trzy miniprocedury	257
Rozwiązywanie problemów — szybkie poprawki	258
Ustawienia startowe — gotowe do użycia	258
Lista kontrolna jakości i bezpieczeństwa — przypnij na stałe	259
Twój kolejny krok	260
Bibliografia	260
20 OBEJŚCIA LIMITÓW I STRATEGIE OSZCZĘDZANIA PRZY KORZYSTANIU Z API	262
Dlaczego istnieją limity zapytań (i dlaczego to działa na Twoją korzyść)	262
Czytaj sygnały: spraw, by HTTP pracowało dla Ciebie	263
Ukształtuj klienta: nie uderzaj w ścianę	264

Najpierw kolejkuje, potem pytaj — architektura kolejki zadań	265
Grupuj, dziel i pamiętaj — trzy dźwignie oszczędności przy API	266
Normalizacja tekstu = ukryte oszczędności tokenów	266
Poziomy jakości i łagodne pogorszenie usług	267
Wybierz właściwy moment — przesunięcie w czasie to oszczędność	268
Wybór dostawcy i modelu — bez bałaganu	268
Obserwowalność, która sama na siebie zarabia	269
Zarządzanie: SLO, budżety błędów i bezpieczne wdrożenia	269
Sprawdzone wzorce, które warto zaadaptować	269
Przewodnik obniżania kosztów — według zwrotu z inwestycji (ROI)	271
Typowe pułapki (i proste rozwiązania)	272
Studium przypadku: od chaosu do spokoju w 48 godzin	272
Mentalność, która wygrywa	273
Bibliografia	273
21 CO DALEJ: SZTUCZNA INTELIGENCJA MULTIMODALNA, DUBBING W CZASIE RZECZYWISTYM I GŁOSOWE NFT	275
Sztuczna inteligencja multimodalna: jeden umysł, wiele zmysłów	275
Dubbing w czasie rzeczywistym zabrzmi naturalnie, a nie jak „dubbing”	277
„Głosowe NFT”: programowalne prawa, a nie kolekcjonerska moda	278
Zaufanie, bezpieczeństwo i pochodzenie w kontekście mowy syntetycznej	280
Wzorce biznesowe, które mają szansę zwyciężyć	281
Zbuduj to: plan działania na najbliższy czas	281
Rejestr ryzyk (z planami łagodzenia)	282
Spojrzenie nieco dalej w przyszłość	283
Bibliografia	283
A GOTOWE SKRYPTY W PYTHONIE I VBA DO MASOWEGO GENEROWANIA GŁOSU	285
Bibliografia	298
B SŁOWNICZEK TERMINÓW I SKRÓTÓW ZWIĄZANYCH Z AI AUDIO	301
Bibliografia	315

2

JAK W ZALEDWIE 15 MINUT STWORZYĆ SWÓJ PIERWSZY GŁOSOWY KOMENTARZ Z WYKORZYSTANIEM TECHNOLOGII SYNTEZY MOWY

„Ludzki głos to najdoskonalszy instrument ze wszystkich — chyba że masz napięty termin, a studio jest już zajęte” — Maya Drew, producentka filmów dokumentalnych.

W tym rozdziale przejdziesz od pustej strony do dopracowanej, gotowej do udostępnienia narracji — i to w czasie jednej przerwy na kawę. Nie będziesz potrzebować ani mikrofonu, ani paneli akustycznych, ani specjalistycznej wiedzy z zakresu inżynierii dźwięku. Wystarczą klawiatura, przeglądarka internetowa oraz bezpłatna wersja narzędzia ElevenLabs.

Wyzwanie, któremu sprostasz w zaledwie 15 minut

Najnowszy model Flash v2.5 firmy ElevenLabs umożliwia przesyłanie mowy z opóźnieniem poniżej 100 milisekund, co oznacza, że możesz odsłuchać wygenerowane zdanie niemal natychmiast po naciśnięciu przycisku *Generate* (generuj). To właśnie ta szybkość stanowi fundament naszej obietnicy: od pustej strony do gotowego do pobrania nagrania audio w zaledwie 15 minut. [1]

Co przygotujesz? Jednominutowy materiał objaśniający, zapisany w formatach MP3 oraz WAV — gotowy do publikacji jako krótki film na YouTube, materiał do Reels, podcastu lub próbki audiobooka. [3]



Zwykły kuchenny minutnik pomoże Ci zachować dyscyplinę czasową, choć równie dobrze sprawdza się tu intuicja — zwłaszcza przy korzystaniu z poniższego przewodnika tempa.

W tabeli 2.1 znajduje się harmonogram pracy.

Tabela 2.1. Rozpisany krok po kroku harmonogram pracy, uporządkowany według kolejnych minut

Znacznik czasu	Działanie	Rezultat
00:00 – 03:00	Rejestracja i logowanie	Otwarty panel użytkownika; przyznane 10 000 bezpłatnych znaków tekstu [3]
03:01 – 06:00	Przesłuchanie dostępnych głosów	Wybrany głos z biblioteki — przypisany do Twojego projektu
06:01 – 09:00	Opracowanie tekstu przyjaznego dla ucha	130 – 150 słów, ocenionych pod kątem przejrzystości i zrozumiałości
09:01 – 12:00	Dodanie znaczników SSML	Wstawione pauzy, akcenty oraz wskazówki reżyserskie
12:01 – 14:00	Generowanie i odsłuch	Wygenerowane nagranie, dopracowane ustawienia za pomocą suwaków
14:01 – 15:00	Eksport i archiwizacja	Zapisane pliki MP3 i WAV; utworzony folder projektu



Przy pierwszym podejściu trzymaj się wyznaczonego tempa — na improwizację przyjdzie czas później.

Minuty 0 – 3 • Zakładanie konta

1. Wejdź na stronę ElevenLabs.io i wybierz opcję *Start for free* (rozpocznij bezpłatnie).
2. Potwierdź adres e-mail — po automatycznym przekierowaniu znajdziesz się w *Dashboard* (panelu użytkownika), gdzie zostanie Ci przyznany miesięczny limit 10 000 znaków.
3. Wskazówka dotycząca bezpieczeństwa — zanim przystąpisz do pracy nad materiałem dla klienta, włącz *Two-factor Auth* (uwierzytelnianie dwuskładnikowe) w zakładce *Settings/Security* (ustawienia/bezpieczeństwo).

Minuty 3 – 6 • Dobór głosu lektorskiego

Wejdź do *Voice Library* (biblioteki głosów) i zastosuj filtr *Style/Narration* (styl/narracja). Najedź kursorem na wybrany głos, aby odsłuchać próbkę — wybierz ten, który najlepiej oddaje charakter Twojej marki, np.: *Ella — Warm Educator* (Ella — ciepła edukatorka).

Skrót dla zaawansowanych. Jeśli dysponujesz już 60-sekundowym nagraniem czystej mowy, przejdź bezpośrednio do zakładki *VoiceLab/Instant Clone*. Model wygeneruje zadowalającą replikę Twojego głosu w niespełna dwie minuty.

Minuty 6 – 9 • Pisanie z myślą o uchu

Tekst ekranowy uwielbia przecinki, lecz mowa — kropki. Staraj się budować zdania krótkie, najlepiej nieprzekraczające 20 wyrazów, zawierające po jednej myśli.

Możesz skorzystać z następującego schematu:

Zacząpka (jedno zdanie) → **Co i dlaczego** (dwa zdania) → **Zachęta do działania** (jedno zdanie).

Gotowy tekst — o objętości 130 – 150 słów — wklej do modułu *Speech Synthesis* (syntezy mowy).

Minuty 9 – 12 • Wzbogacanie wypowiedzi za pomocą języka SSML

Platforma ElevenLabs obsługuje wybrane elementy języka znaczników SSML (ang. *Speech Synthesis Markup Language* — język znaczników syntezy mowy), opracowanego przez konsorcjum W3C. Na co dzień będziesz korzystać z trzech podstawowych znaczników [2]:

```
<break time="400ms"/>
<emphasis level="strong">
<prosody rate="85%">
```

Znacznik `<break>` wprowadza pauzę — najczęściej wystarcza 300 – 500 milisekund, by oddać funkcję przecinka w mowie.

Znacznik `<emphasis>` pozwala zaakcentować kluczowe słowa, nie powodując przy tym nienaturalnego podwyższenia tonu.

Zmiana tempa wypowiedzi (`<prosody rate>`) spowalnia trudniejsze fragmenty — doskonale sprawdza się przy liczbach, adresach internetowych czy nazwach własnych.



Stosuj znaczniki z umiarem — ich nadmiar sprawia, że wypowiedź brzmi sztucznie. Badania akademickie dowodzą, że słuchacze preferują subtelne, domyślne sygnały emocjonalne, zamiast tych wyrażanych wprost. [1]

Minuty 12 – 14 • Generowanie, odsłuch, dopracowanie

Wybierz odpowiedni model syntezy: *Eleven v3*, jeśli zależy Ci na głębi i naturalności brzmienia, lub *Flash v2.5*, gdy priorytetem jest szybkość działania.

Suwak stabilności (*Stability*): wartość 0 odpowiada aktorowi improwizującemu, 100 — prezyterowi wiadomości. Zalecany punkt wyjścia: około 70.

Naciśnij *Generate* (generuj). Twoje nagranie — liczące około 150 słów — pojawi się w ciągu 2 – 4 sekund.

Nie brzmi idealnie? Delikatnie zwiększ parametr *Style Exaggeration* (wyolbrzymienie stylu), by dodać wypowiedzi wyrazistości, lub wstaw znacznik <break> przed dłuższym członem zdania, by poprawić rytmikę.

Minuty 14 – 15 • Eksport i uporządkowanie materiału

Plik MP3 o przepływności 128 kb/s — gotowy do publikacji w mediach społecznościowych (około 550 KB na 30 sekund nagrania).

Plik WAV o parametrach 44,1 kHz/16 bitów — idealny do profesjonalnego masteringu bez strat jakości (około 5 MB na 30 sekund).

W panelu *Projects* (projekty) wybierz opcję *New Folder* (nowy folder) i nazwij swój folder *Launch Explainer – sierpień 2025*. Wszystkie przyszłe wersje zostaną uporządkowane w jednym miejscu.

Gratulacje! Udało Ci się zmieścić w wyznaczonym limicie 15 minut.

Warsztat doskonalący: od tweeta do minutowego podcastu

Wklej dowolny edukacyjny wątek z Twittera (około 400 znaków) do edytora tekstu. Każde zdanie oddziel znacznikiem <break time="300ms"/>, a następnie wybierz dwa głosy o wyrażnie odmiennym charakterze. Przypisz *Voice 1* (głos 1) do akapitów nieparzystych, a *Voice 2* (głos 2) do parzystych. Po wygenerowaniu nagrania zastosuj w programie Audacity efekt przenikania (*cross-fade*), dodając muzykę w tle o poziomie głośności -18 LUFS. [4]



Efekt końcowy przypomina zwięzły program prowadzony przez duet lektorów — idealny do nowej ścieżki audiobooków z narracją AI, rozwijanej przez Spotify.

Najczęściej napotymane trudności oraz sprawdzone metody ich rozwiązania

W tabeli 2.2 zebrałem typowe przeszkody i sposoby ich przewyciężenia.

Tabela 2.2. Typowe przeszkody, ich przyczyna i propozycje rozwiązania

Problem	Przyczyna	Rozwiązanie
Niewyraźna artykulacja sylab	Brak przecinków w tekście	Wstaw znacznik <break> lub podziel zdanie na krótsze.
Komunikat „Rate-limit exceeded” (przekroczono limit żądań)	Zbyt wiele zapytań na minutę w wersji bezpłatnej	Odczekaj 60 sekund lub przejdź na plan Starter.
Błędna wymowa specjalistycznego terminu	Nieprecyzyjne przekształcenie grafemu na fonem	Zapisz wyraz fonetycznie lub użyj zapisu w alfabecie IPA.
Syczenie przy miksowaniu z muzyką	Niezgodność częstotliwości próbkowania	Przekonwertuj oba pliki na format WAV o częstotliwości próbkowania 44,1 kHz.



Alfabet IPA (międzynarodowy alfabet fonetyczny) to standardowy system symboli stworzony przez Międzynarodowe Towarzystwo Fonetyczne, który pozwala na precyzyjne zapisywanie dźwięków każdego języka na świecie, służąc do analizy i nauki wymowy w nauce języków obcych. Jest on powszechnie używany w lingwistyce i umożliwia jednoznaczne przedstawienie fonemów, czyli najmniejszych jednostek dźwiękowych mowy, pomagając przewyciężyć niespójności w tradycyjnym piśmie. [5]

Najważniejsze wnioski

- **Szybkość to atut.** Dynamiczne iteracje sprzyjają eksperymentom — zamiast jeden raz spróbuj przeczytać tekst trzykrotnie.
- **Dobrze napisany tekst przewyższa najbardziej wyszukane znaczniki.** Najpierw pisz w sposób naturalny, konwersacyjny; oznaczenia techniczne dodawaj tylko wtedy, gdy rozwiązują konkretny problem. [6] [7]
- **Format ma znaczenie.** Wybierz MP3, jeśli zależy Ci na szybkim opublikowaniu materiału; WAV zapewni natomiast większą swobodę przy edycji.

- **Porządek w projekcie to oszczędność czasu.** Nazywaj foldery tak, jakby jutro miał je otworzyć Twój współpracownik.

W kolejnym etapie przejdziemy od narracji jednoosobowej do dialogu wielogłosowego (ang. *multi-voice dialogue*), zgłębiając techniki łączenia emocji, obsady głosowej oraz automatyzacji scenariusza.

Bibliografia

- [1] Z. Mu, W. Xihong, Y. Kai, *Review of end-to-end speech synthesis technology based on deep learning*, „Digital Communications and Networks” 2021, 7, no. 4, s. 435 – 455.
- [2] ElevenLabs, *Text to Speech | Documentation*, 2025, <https://elevenlabs.io/docs/capabilities/text-to-speech> [dostęp: 7 listopada 2005].
- [3] ElevenLabs, *Pricing for creators and businesses of all sizes*, 2025, <https://elevenlabs.io/pricing> [dostęp: 7 listopada 2005].
- [4] J. Peters, *Spotify is making it easier to release audiobooks narrated by AI*, The Verge, 5 marca 2025, <https://www.theverge.com/news/616355/spotify-audiobooks-elevenlabs-ai-narration> [dostęp: 7 listopada 2005].
- [5] J. Weatherbed, *ElevenLabs’ AI Reader app can now narrate text in 32 languages*, The Verge, 23 sierpnia 2024, <https://www.theverge.com/2024/8/23/24226643/elevenlabs-ai-reader-app-32-languages-global-release> [dostęp: 7 listopada 2005].
- [6] J. Kim et al., *Glow-TTS: A generative flow for text-to-speech via monotonic alignment search* [w:] *Advances in Neural Information Processing Systems* 33, 2020, s. 8069 – 8082.
- [7] B. Katz, *Mastering Audio. The art and the science*, 3rd ed., Focal Press, 2015, s. 134 – 136.

PROGRAM PARTNERSKI

— GRUPY HELION —

- 
1. ZAREJESTRUJ SIĘ
 2. PREZENTUJ KSIĄŻKI
 3. ZBIERAJ PROWIZJĘ

Zmień swoją stronę WWW w działający bankomat!

Dowiedz się więcej i dołącz już dzisiaj!

<http://program-partnerski.helion.pl>

GRUPA
Helion

ElevenLabs

To do mnie przemawia

Wśród rozmaitych zastosowań sztucznej inteligencji znajdują się możliwości klonowania i generowania głosu. Obecnie na rynku dostępnych jest kilka narzędzi służących do tych celów — jedno z nich to wciąż zyskująca na popularności platforma ElevenLabs. Stworzył ją zespół pod przewodnictwem dwóch polskich inżynierów: Piotra Dąbkowskiego, który wcześniej pracował dla Google, i Mateusza Staniszewskiego, byłego stratega wdrażania w Palantir Technologies. Inspiracją do powstania ich firmy ElevenLabs była chęć kreowania bardziej realistycznych dubbingów filmowych.

Jeśli chcesz się nauczyć klonować głosy w taki sposób, by były nieodróżnialne od prawdziwych — odpowiednio modulowane, o oczekiwanej barwie i intonacji, przekazujące emocje — ta książka jest dla Ciebie. Obsługa narzędzi do generowania głosów AI nie jest trudna, a dzięki temu poradnikowi opanujesz ją błyskawicznie.

Z książki dowiesz się między innymi:

- Dlaczego spośród dostępnych syntezytorów mowy warto wybrać ElevenLabs
- Jak szybko stworzyć swój pierwszy głosowy komentarz
- W jaki sposób promptować, by uzyskać naturalnie brzmiące rezultaty
- Co o klonowaniu i generowaniu głosów mówi prawo
- Jak przygotowywać nagrania wysokiej jakości — w tym na potrzeby radia i w różnych językach
- Jak generować dialogi, podcasty i audiobooki
- W jaki sposób można zarabiać dzięki syntezie głosu opartej na AI

	KOD KORZYŚCI <i>Sięgnij po więcej!</i> 
 helion.pl	ISBN 978-83-289-3488-7  9 788328 934887
 HELION S.A. ul. Kościuszki 1c 44-100 Gliwice tel.: 32 230 98 63 helion@helion.pl	Cena: 89,00 zł