

Bouwen in Dify: van opname naar verslag

Handleiding 5, bij use case UC-2026-009 AI-Transcriptietool

Experimenteerruimte ICT SOML · LeX Consultancy B.V. · 20 september 2026



Je bouwt een workflow die een geluidsopname van een overleg omzet in tekst en daar een verslag van maakt: samenvatting, besluiten, actiepunten met eigenaar en datum, en wat is doorgeschoven. Aan het eind heb je een webadres waar een collega een mp3 uploadt en binnen twee minuten het verslag en het volledige transcript terugkrijgt.

Ten opzichte van handleiding 1 en 3 zijn twee dingen nieuw: een **bestand als invoer** in plaats van tekst, en een **tool-node** (Speech To Text) in plaats van alleen LLM-nodes. De rest ken je.

Deze handleiding volgt de cloudversie van Dify zoals die er op 20 september 2026 uitziet. Dify verandert snel: kijkt je scherm af, volg dan wat je ziet en niet wat hier staat.

Zo lees je deze handleiding

Vorm	Betekenis
1. 2. 3.	Wat je doet: klikken, kiezen, openen
Vet	Een knop, menu of veld zoals het op je scherm staat
Tabel <i>Wat je intypt</i>	Waarden die je in een formulierveld zet
Grijs kader <i>Intypen</i>	Tekst die je letterlijk overtypt of plakt
Oranje kader	Een valkuil die je anders zelf ontdekt
Groen kader <i>Zelf kiezen</i>	Een plek waar je ons voorbeeld kunt vervangen door je eigen materiaal, stijl of prompt

De ontwerpkaart van deze app

Voordat we gingen klikken, is deze kaart ingevuld (de lege kaart staat in de bijlage Ontwerpkaart). Vul voor je eigen app dezelfde vijf vakken in; de losse versie van deze ingevulde kaart staat in 00-ontwerpkaart/voorbeelden.

1. Voor wie	2. Wat gaat erin
De notulist van een sectieoverleg die het verslag altijd twee weken later pas af heeft. Heeft een mp3 van drie kwartier en een kwartier tijd.	Eén audiobestand (mp3, m4a of wav, maximaal 5 MB) via een bestandsveld. Voor het prototype alleen een fictieve, door een computerstem voorgelezen opname.

3. Wat komt eruit	4. Waar moet hij zich aan houden
Een verslag met vijf kopjes (Overleg, Samenvatting, Besluiten, Actiepunten met eigenaar en datum, Doorgeschoven) en daaronder het volledige transcript. Onzekere namen met een vraagteken.	Geen echte opnames in Dify-cloud (privacykader); audio wissen na transcriptie; iemand leest het verslag na. Tool: Speech To Text met een eigen OpenAI-sleutel.

5. Waaraan zie je dat het werkt

Test	Wat gaat erin	Wat moet eruit komen
1	De fictieve opname van drie minuten	Verslag met vijf kopjes, drie actiepunten met eigenaar
2	Dezelfde opname als 128 kbps (te groot)	Foutmelding over de bestandsgrootte; verkleinen naar 32 kbps
3 (de lastige)	Een foto in plaats van audio	Het invoerveld weigert het bestand; alleen Audio staat aan

Lees dit eerst: wat mag wel en niet

Deze handleiding hoort bij een use case die door de Experimenteerruimte ICT en de Functionaris Gegevensbescherming is beoordeeld: **voorwaardelijk akkoord, risico middel**. Een stem is een persoonsgegeven. Daarom gelden deze afspraken, en die staan niet ter discussie:

Wat	Afspraak
Opnames van echte mensen in Dify-cloud	Niet. Dify-cloud valt niet onder de verwerkersovereenkomst van SOML. Deze handleiding werkt met een fictieve, door een computer voorgelezen opname. Gebruik die, of maak zelf zo'n opname (bijlage A).
Echt gebruik	Alleen via de Gemini API in het centrale Google Cloud-project van SOML (Vertex AI, verwerking in de EU). Regel dat via e-ict@soml.nl. Dify kan daar ook op draaien, maar dat is een aparte inrichting die buiten deze handleiding valt.
Opnemen van een overleg	Vooraf expliciete toestemming van alle deelnemers. Structureel opnemen: medezeggenschapsraad informeren.
Bewaren	Audio verwijderen zodra het transcript er is. Dify bewaart geüploade bestanden en resultaten; zie "Beheer" achterin.
Leerlingen	Het scenario met spreekvaardigheidstoetsen mag pas na een DPIA en na het informeren van ouders. Deze handleiding gaat alleen over overleggen van medewerkers.
Toegankelijkheid	Als de tool voor leerlingen wordt ingezet: voor alle leerlingen, niet alleen voor leerlingen met een beperking.

Kort gezegd: wat je hier bouwt is een werkend prototype om te leren hoe het gaat. De stap naar echt gebruik is een aparte beslissing met een aparte omgeving.

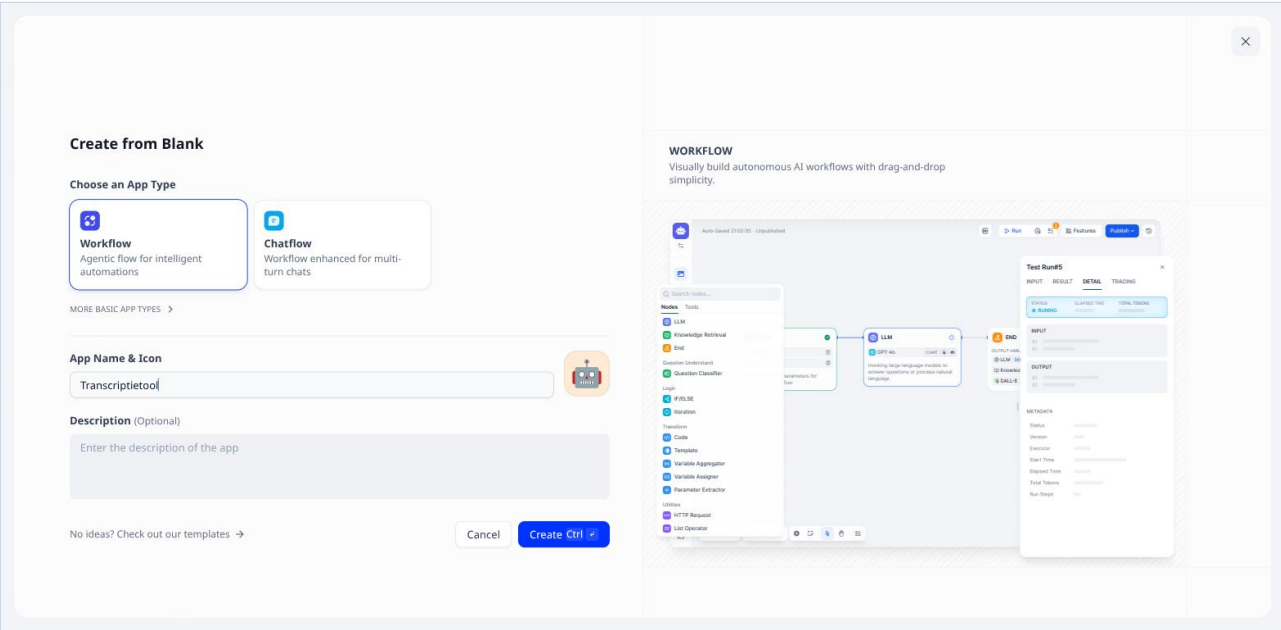
Vorbereiding

- Een Dify-account met een eigen OpenAI-sleutel, ingesteld op **Gebruiksprioriteit: API-sleutel** (Integraties → Model Provider → OpenAI → Config). Zonder eigen sleutel werkt de Speech To Text-tool niet: de gratis credits van Dify dekken alleen tekstmodellen.
- De testopname `sectieoverleg-opname-klein.mp3` uit de map van deze handleiding: een fictief sectieoverleg Nederlands van drie minuten, drie sprekers, voorgelezen door een computerstem.
- Kosten per run via je eigen sleutel: transcriptie ongeveer 2 cent, verslag minder dan 1 cent.

Stap 1. Maak de workflow

1. Ga naar **Studio**, klik op **Create** en kies **Create from Blank**.
2. Klik op **Workflow**.
3. Vul de naam in.

Veld op het scherm	Wat je intypt
App Name & Icon	Transcriptietool



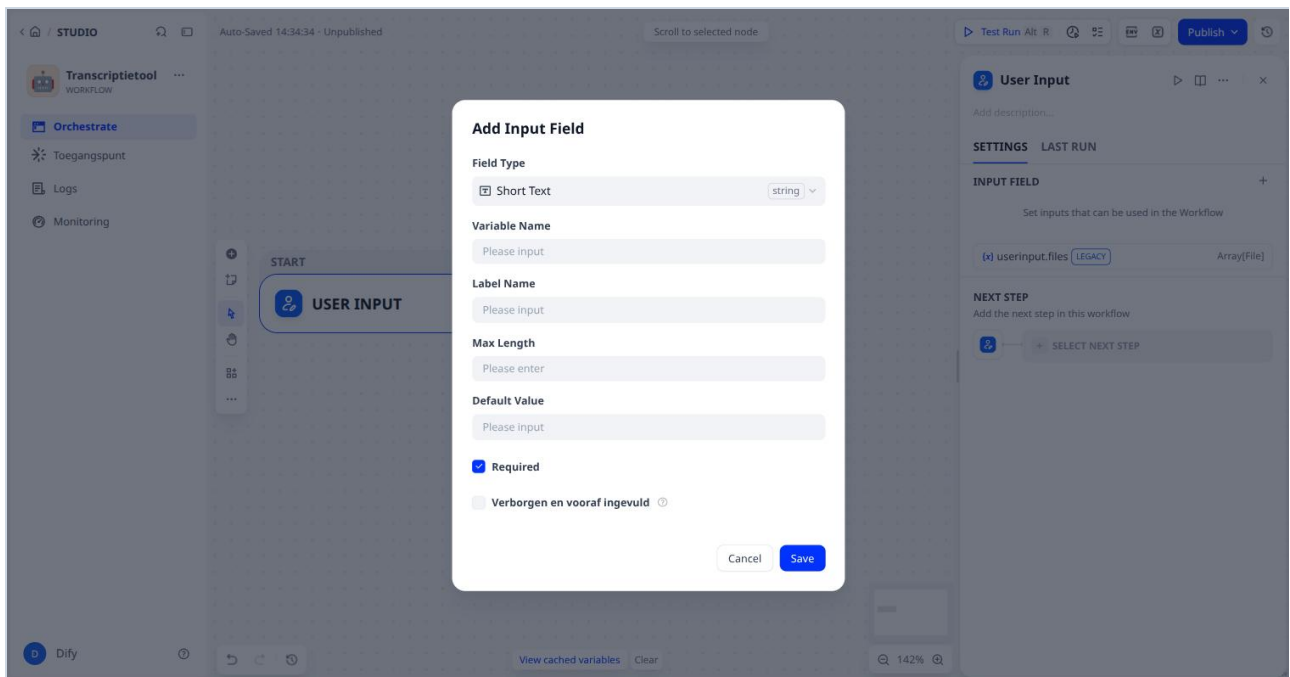
Workflow gekozen, naam ingevuld

4. Klik op **Create**.

Een sandbox-account mag vijf apps hebben, en agents in het menu Agents tellen mee. Zie je “Upgrade to create more apps”, verwijder dan eerst een oude app via de drie puntjes op de kaart in Studio.

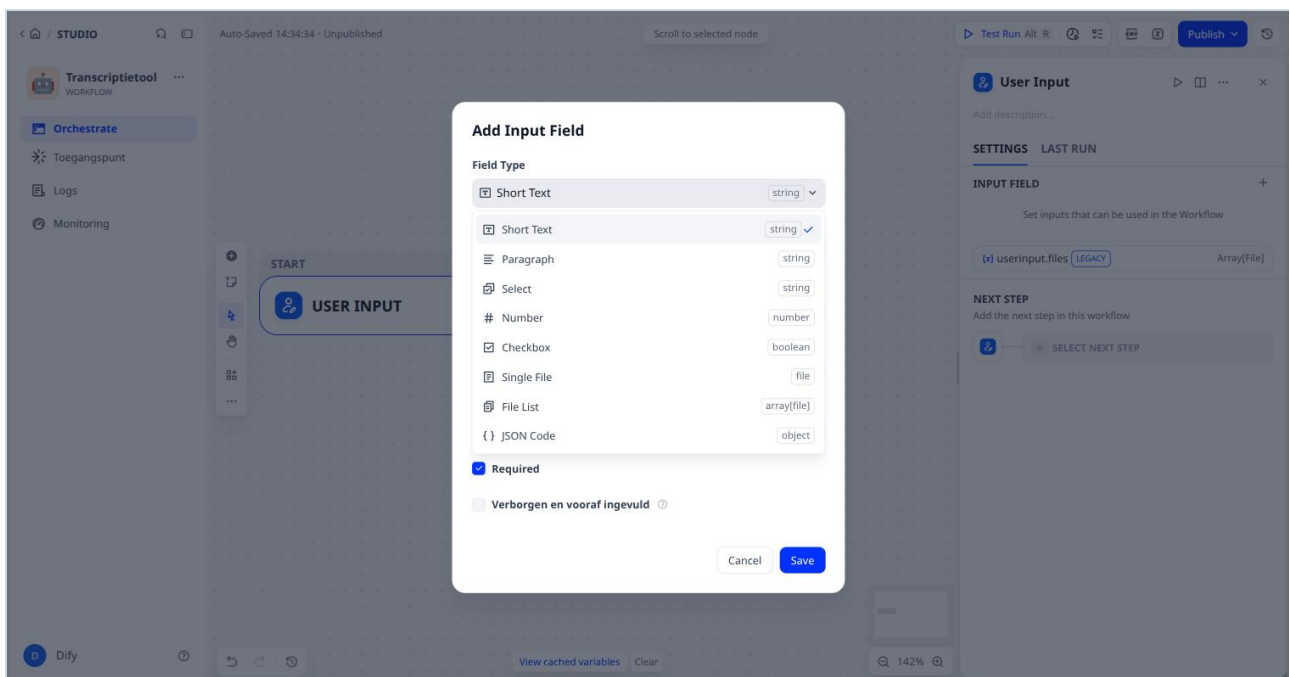
Stap 2. Maak een invoerveld voor het audiobestand

1. Klik in het rechterpaneel op **User Input**.
2. Klik op het plusje bij **INPUT FIELD**.



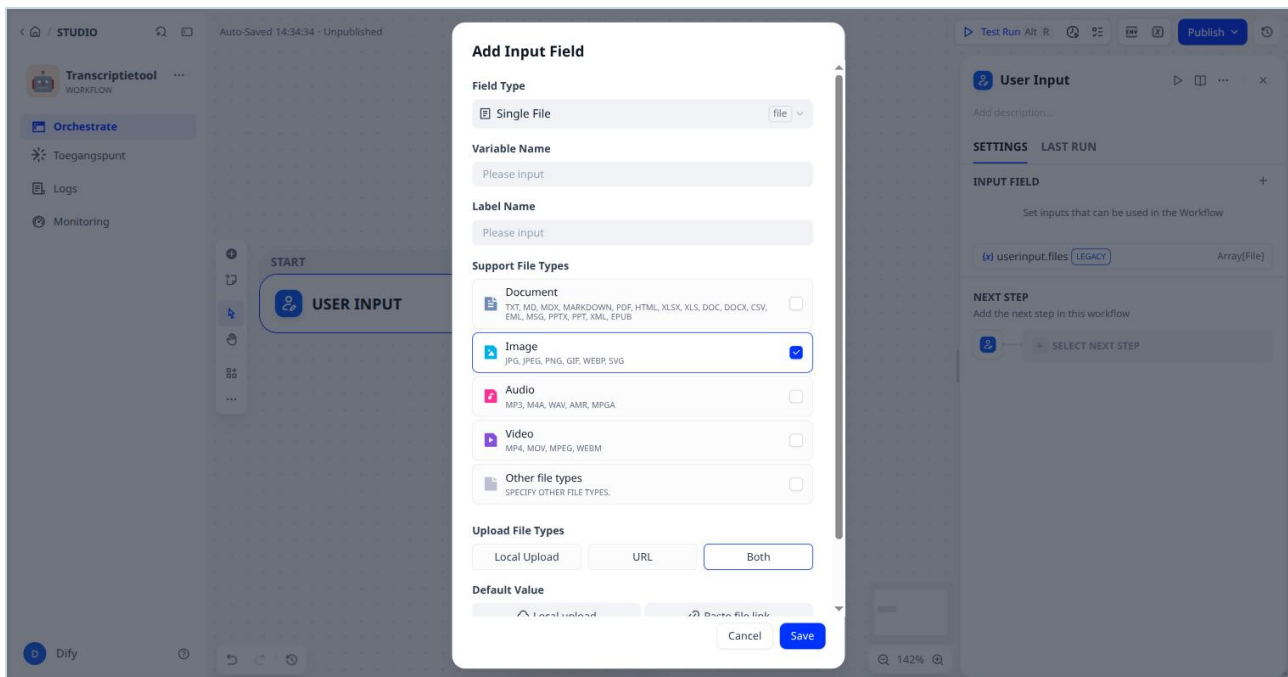
Het dialoogvenster Add Input Field

3. Klap **Field Type** open. Je ziet de soorten velden.



De veldtypen, met Single File en File List

4. Kies **Single File**. Het venster wordt langer: er verschijnen bestandstypen.

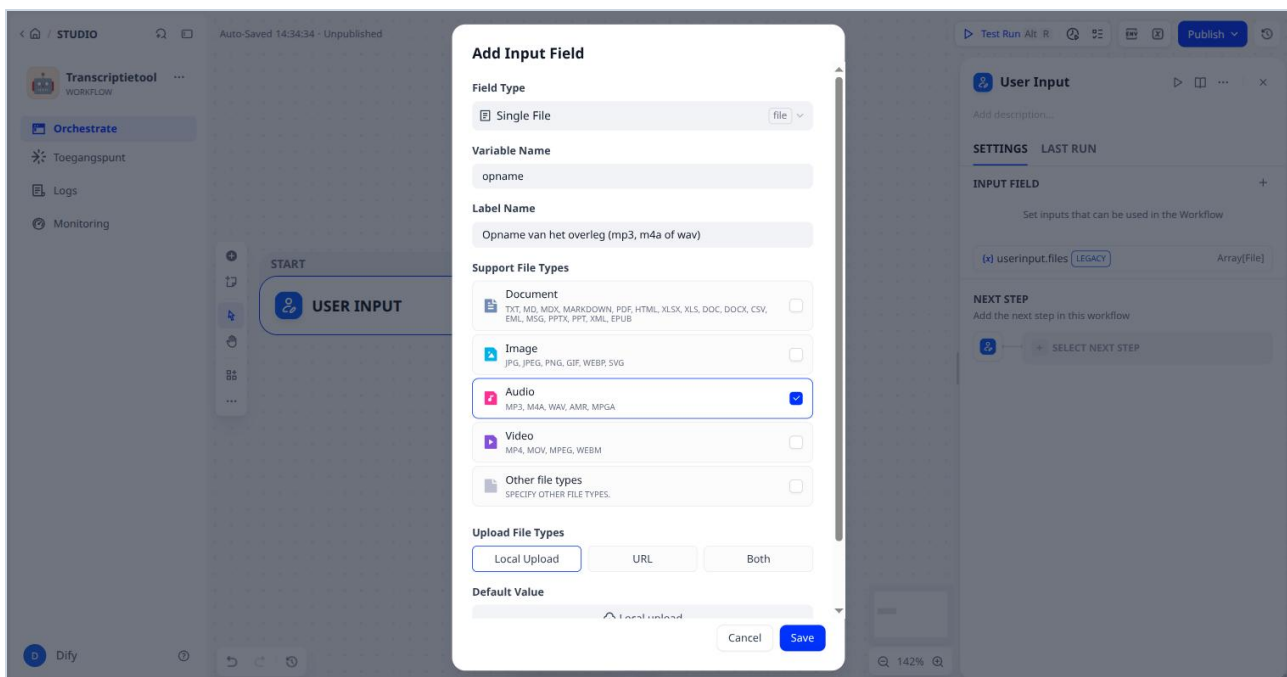


Single File: bestandstypen kiezen

5. Vul de velden in.

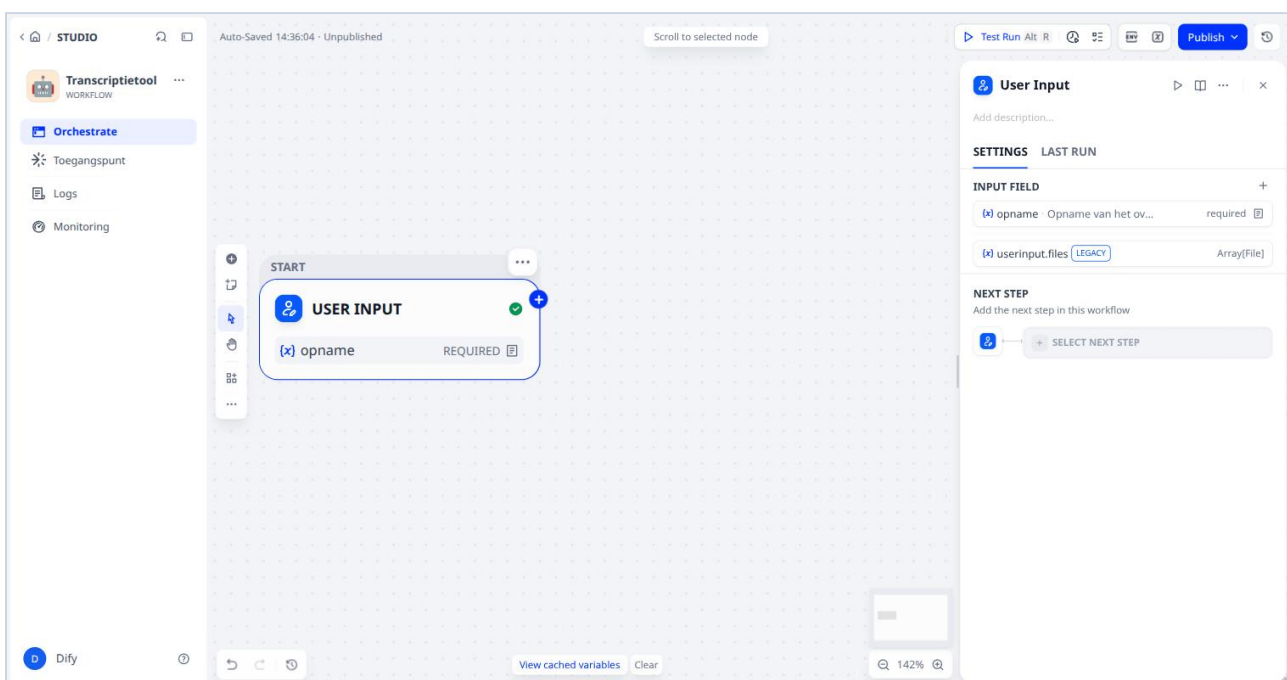
Veld op het scherm	Wat je intypt of kiest
Variable Name	opname
Label Name	Opname van het overleg (mp3, m4a of wav)
Support File Types	Image uitzetten, Audio aanzetten
Upload File Types	Local Upload

Image staat standaard aan. Zet hem uit, anders kan iemand straks een foto uploaden en loopt de workflow vast op de transcriptiestap.



Veld opname, alleen Audio aangevinkt

6. Klik op **Save**.

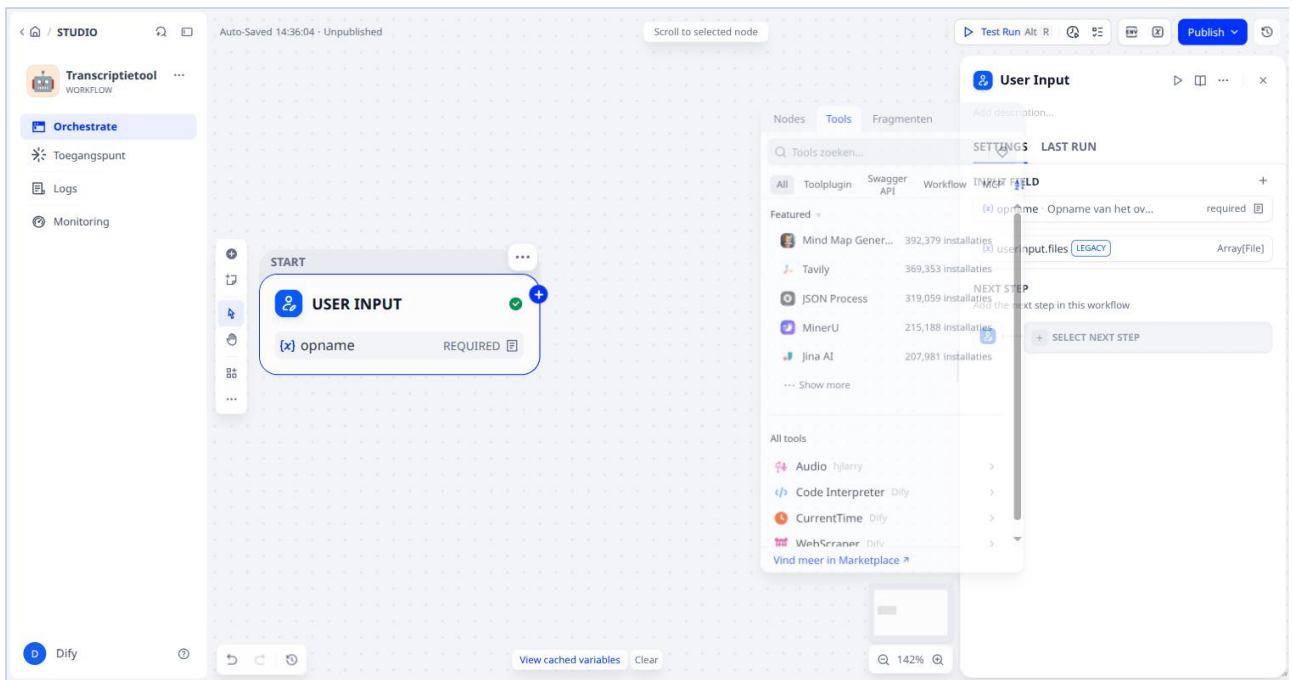


Het invoerveld staat in de node

Stap 3. Voeg de transcriptiestap toe

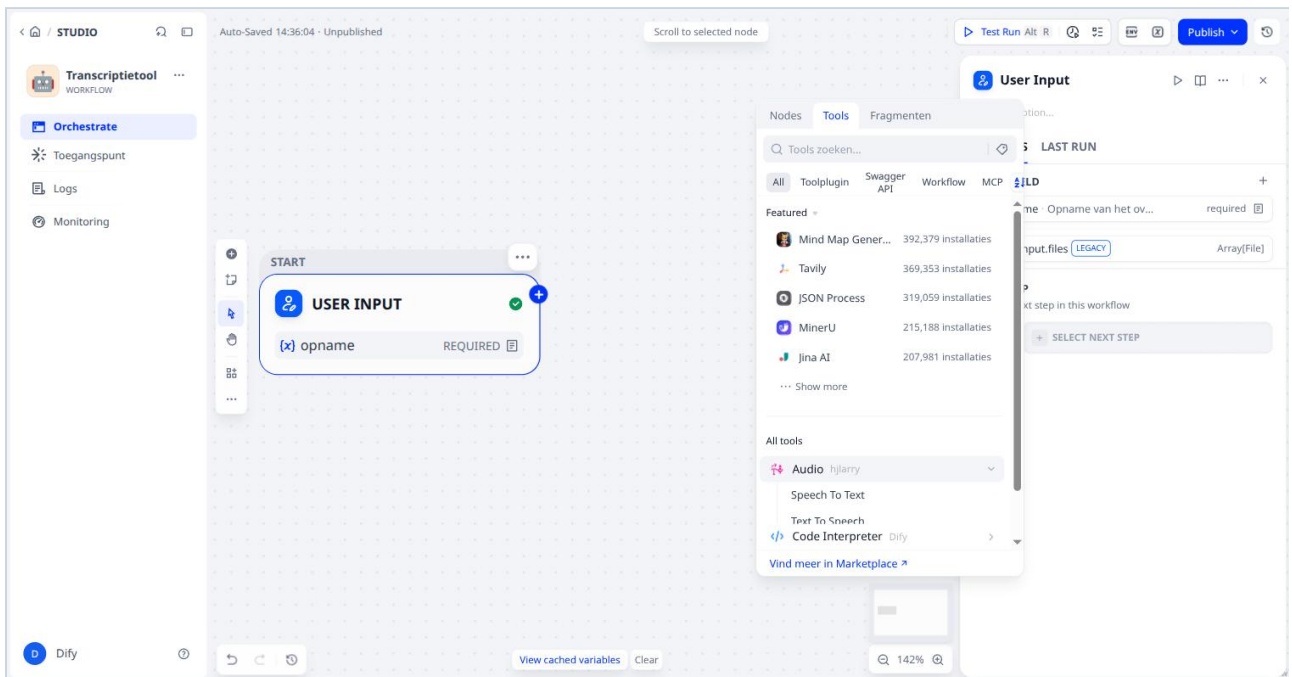
Dit is de nieuwe bouwsteen: geen LLM-node, maar een **tool**. Dify heeft een ingebouwde Audio-tool die spraak naar tekst omzet via het spraakmodel van je aanbieder.

1. Klik onder **NEXT STEP** op **SELECT NEXT STEP**.
2. Klik bovenin het lijstje op het tabblad **Tools**.



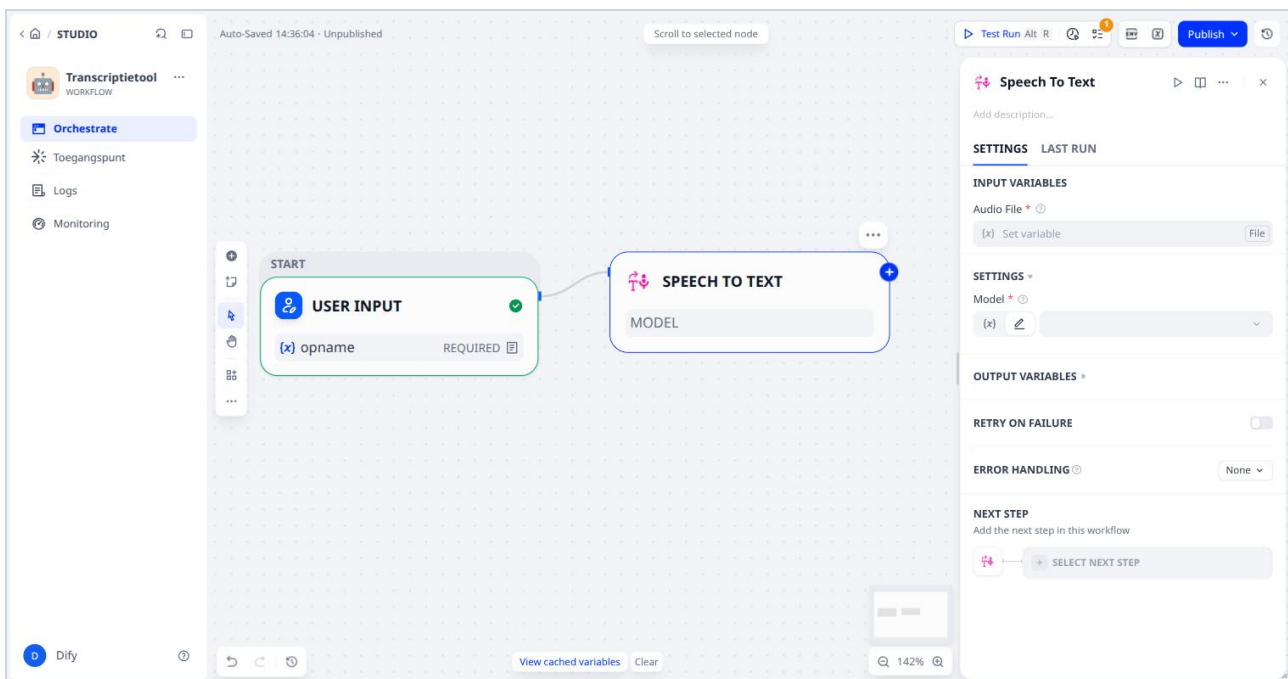
Het tabblad Tools

3. Klik onder **All tools** op **Audio**. Er klappen twee opties uit.



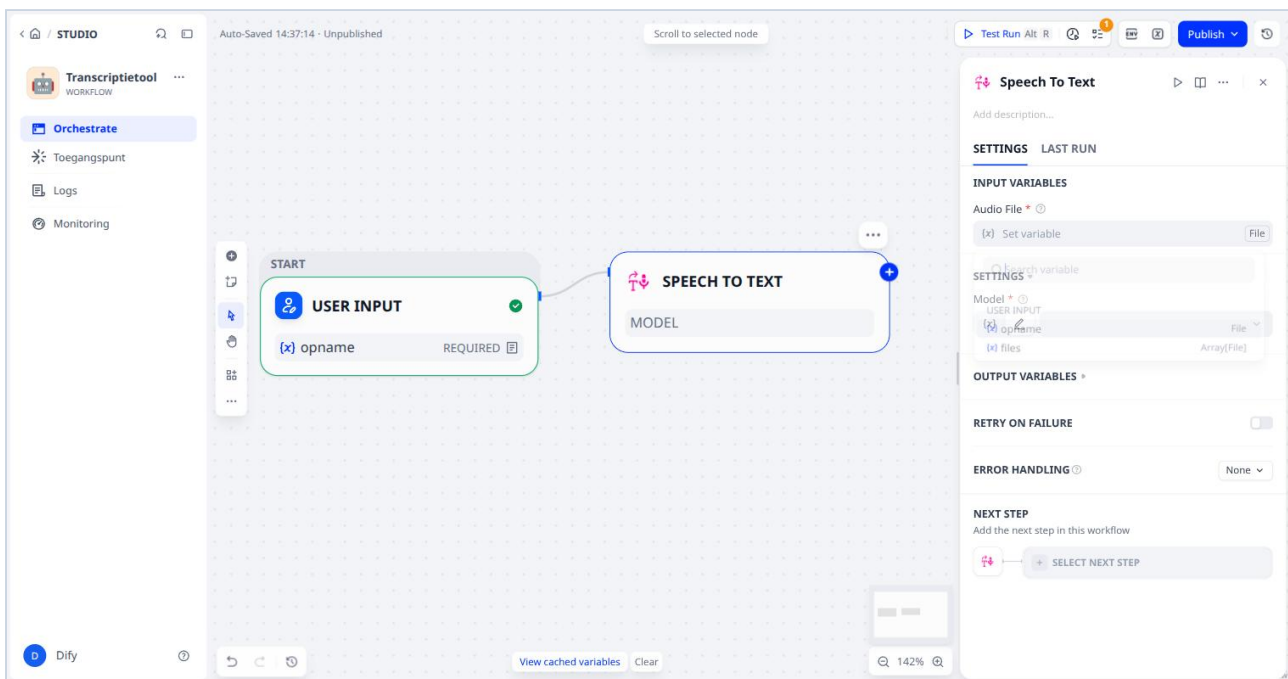
Audio: Speech To Text en Text To Speech

4. Kies **Speech To Text**.



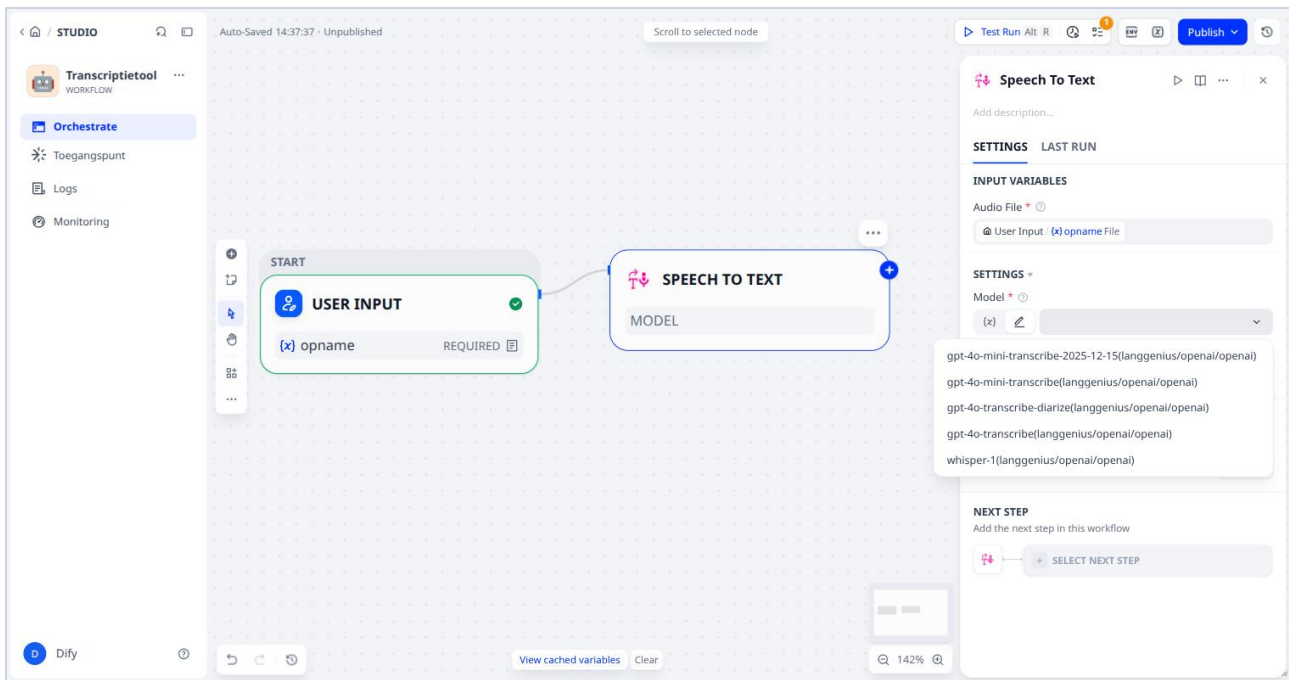
De lege Speech To Text-node

5. Klik bij **Audio File** op **Set variable** en kies onder USER INPUT de variabele **opname**.



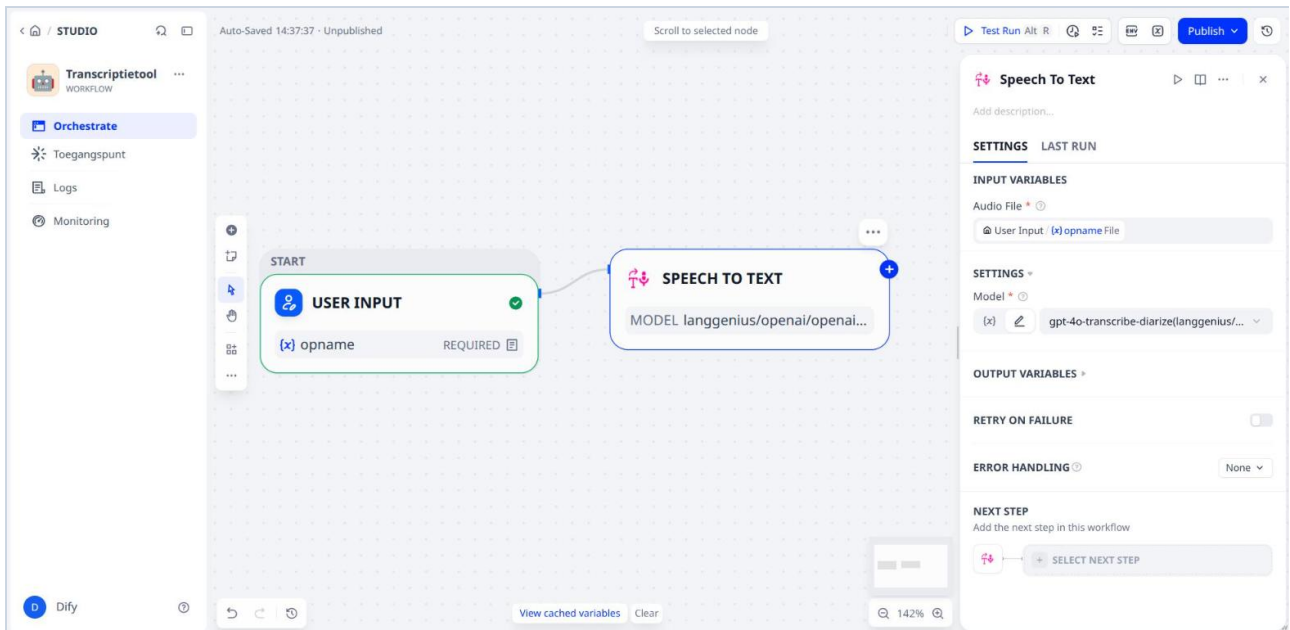
De opname koppelen

6. Klik bij **Model** op de keuzelijst. Je ziet de spraakmodellen die via je sleutel beschikbaar zijn.



De spraakmodellen

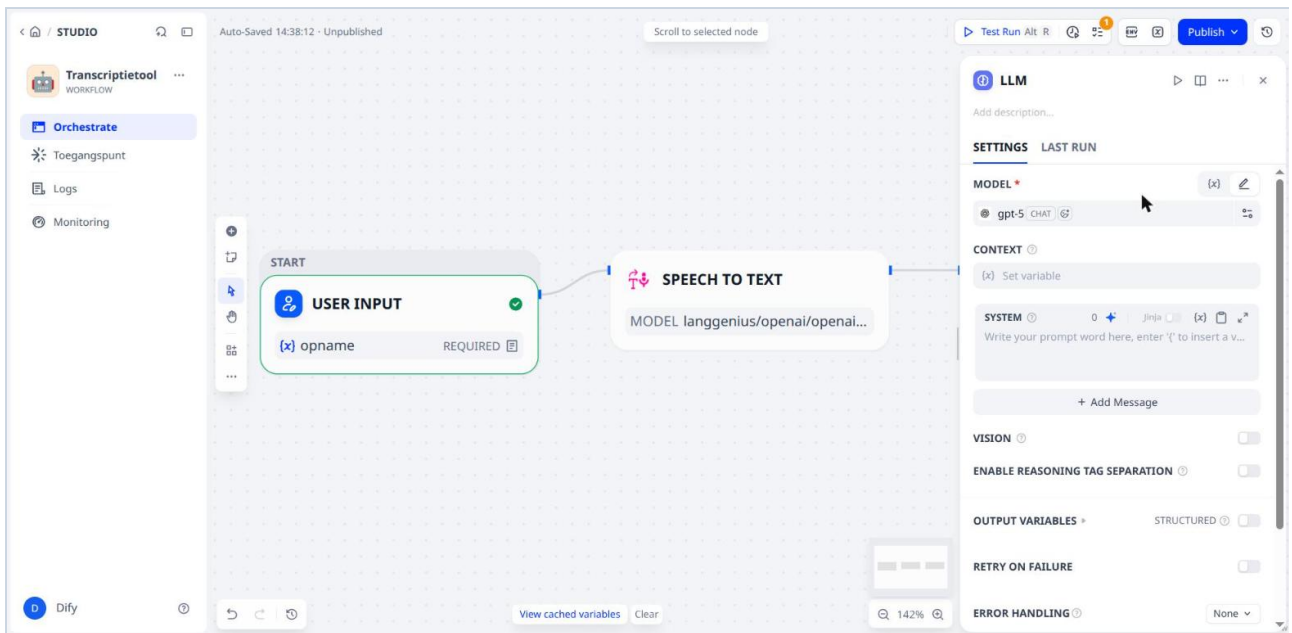
7. Kies **gpt-4o-transcribe-diarize**. "Diarize" betekent dat het model sprekers uit elkaar houdt; dat helpt bij een overleg. Wil je goedkoper, kies dan **whisper-1**.



Speech To Text ingesteld

Stap 4. Voeg de verslagstap toe

1. Klik op **SELECT NEXT STEP** en kies onder **Nodes** de node **LLM**.



De lege LLM-node

2. Klik op de modelnaam **gpt-5** en kies **gpt-5-mini**.
3. Klik in het veld **SYSTEM** en typ de instructie over.

INTYPEN

Begin je antwoord met de regel # VERSLAG en daarna een lege regel.

Je maakt een verslag van een overleg op een school, op basis van het transcript hieronder. Het transcript komt van automatische spraakherkenning en kan fouten bevatten in namen en getallen; schrijf dan op wat het meest waarschijnlijk is en zet een vraagteken tussen haakjes achter wat je niet zeker weet.

Schrijf in het Nederlands, zakelijk en kort. Gebruik precies deze indeling met deze kopjes:

Overleg

Een regel: welk overleg, datum, aanwezigen (als dat uit het transcript blijkt).

Samenvatting

Maximaal vijf zinnen over wat er is besproken.

Besluiten

Genummerde lijst. Alleen wat echt is besloten, geen voornemens.

Actiepunten

Tabel met kolommen: Wat, Wie, Wanneer. Staat er geen datum, schrijf dan "niet genoemd".

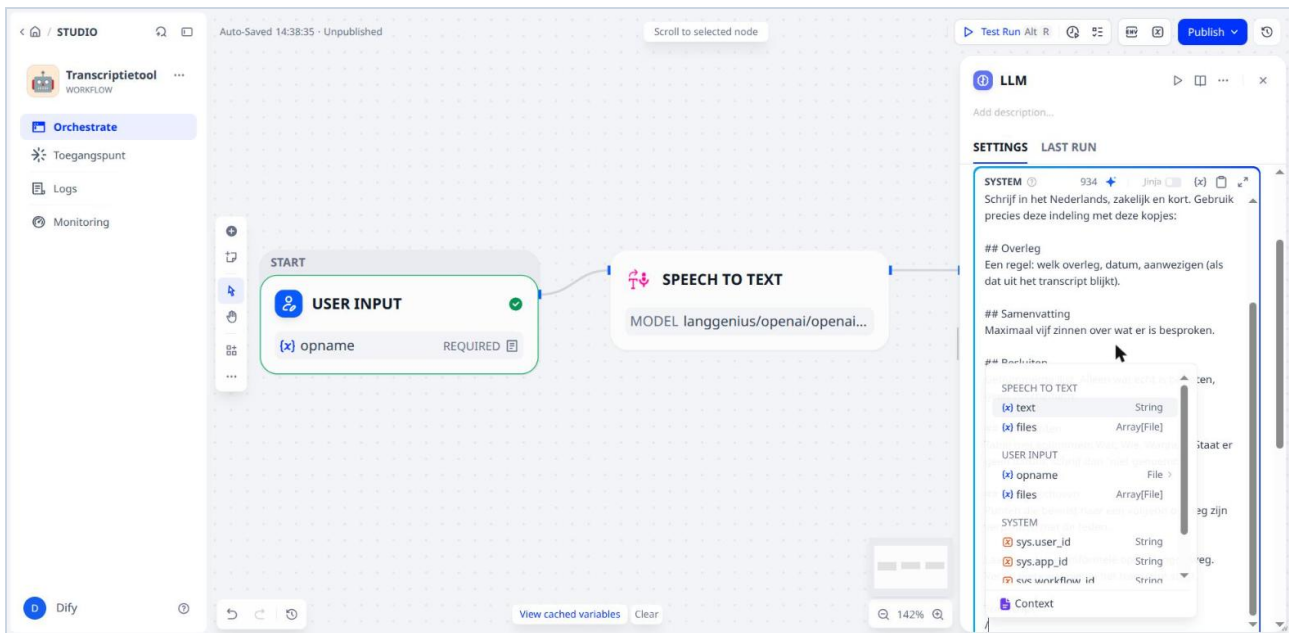
Doorgeschoven

Punten die bewust naar een volgend overleg zijn verplaatst, met de reden.

Laat zijpaden en informele opmerkingen weg. Verzin niets wat niet in het transcript staat.

Transcript:

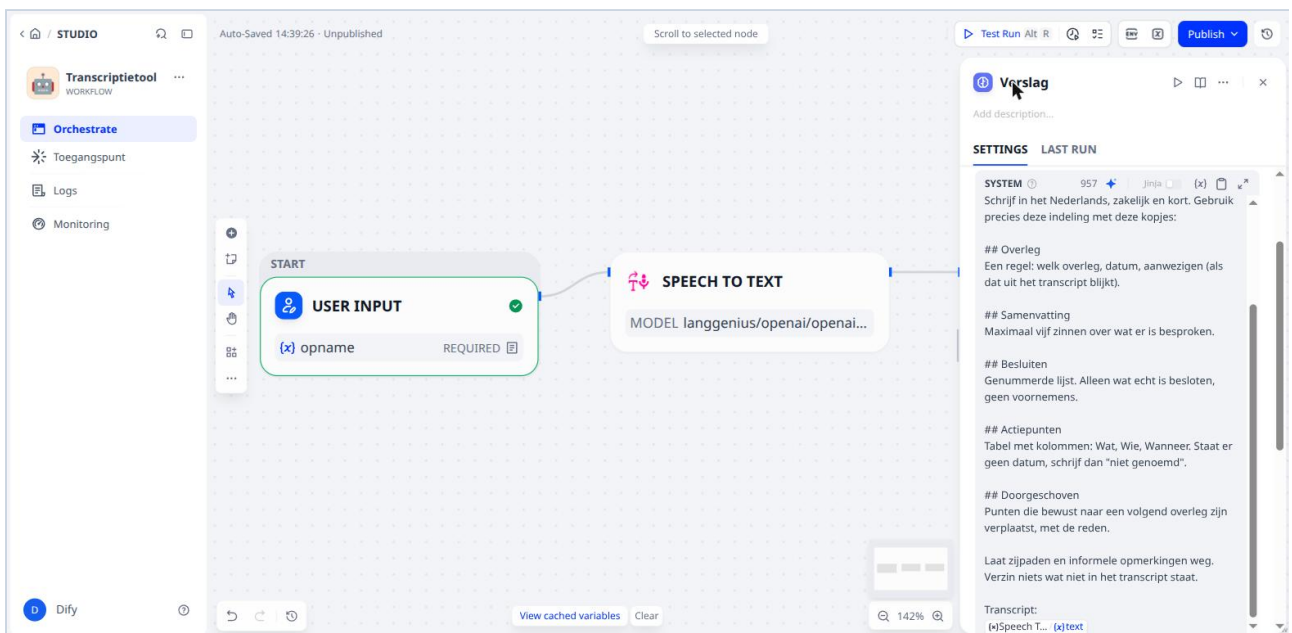
4. Typ direct na "Transcript:" een schuine streep / . In de lijst staat nu bovenaan **SPEECH TO TEXT** met de variabele **text**.



De variabelenlijst met de uitvoer van Speech To Text

5. Kies **text** onder SPEECH TO TEXT.

6. Klik bovenin het paneel op de naam **LLM**, vervang die door **Verslag** en druk op Enter.



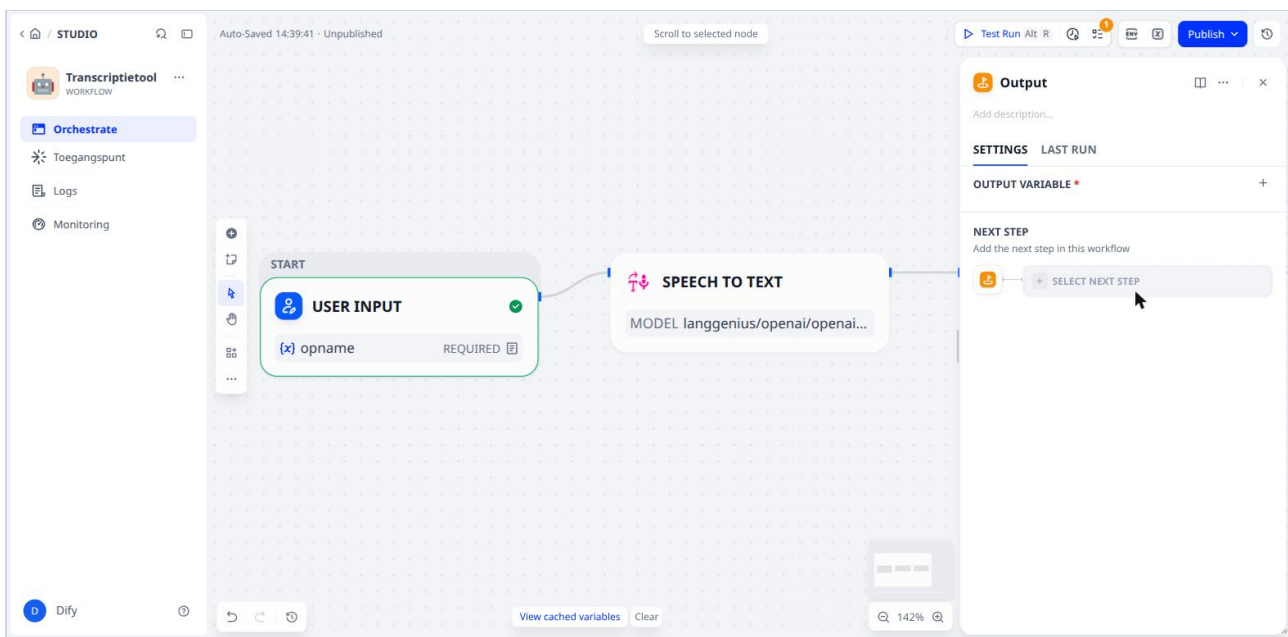
De Verslag-node, compleet

Zelf kiezen: de indeling van het verslag. De kopjes in de prompt zijn de indeling die een sectieoverleg vaak heeft. Een MR-vergadering wil misschien “Adviezen en instemming”, een leerlingbespreking wil per leerling een blokje (maar dat is het scenario dat pas na een DPIA mag), een teamdag wil “Afspraken” en “Parkeerplaats”. Verander alleen de kopjes en hun uitleg; de regels erboven (fouten in spraakherkenning, niets verzinnen) laat je staan. Wil je twee versies, bijvoorbeeld een lang verslag en een bericht van vijf regels voor het team? Voeg een tweede LLM-node toe met dezelfde variabele **text** als invoer.

De regel “zet een vraagteken tussen haakjes achter wat je niet zeker weet” is geen versiering. In onze test hoorde de spraakherkenning “Karin” de ene keer als “Karen” en de andere keer als “Karin”, en werd “Sectieoverleg Nederlands” een keer “Secr. Overleg Nederland”. Het verslag zette daar netjes “(Karen?)” bij. Een mens moet dat nog nakijken; de tool maakt zichtbaar waar.

Stap 5. Bepaal wat eruit komt

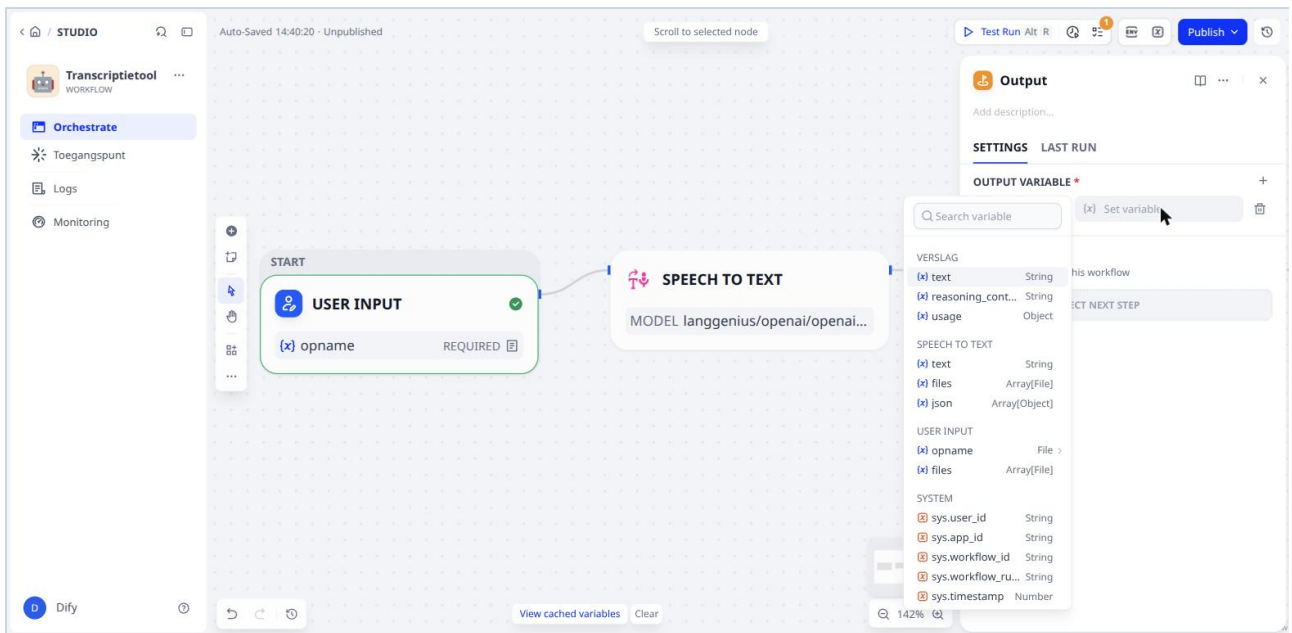
1. Klik op **SELECT NEXT STEP** en kies **Output**.



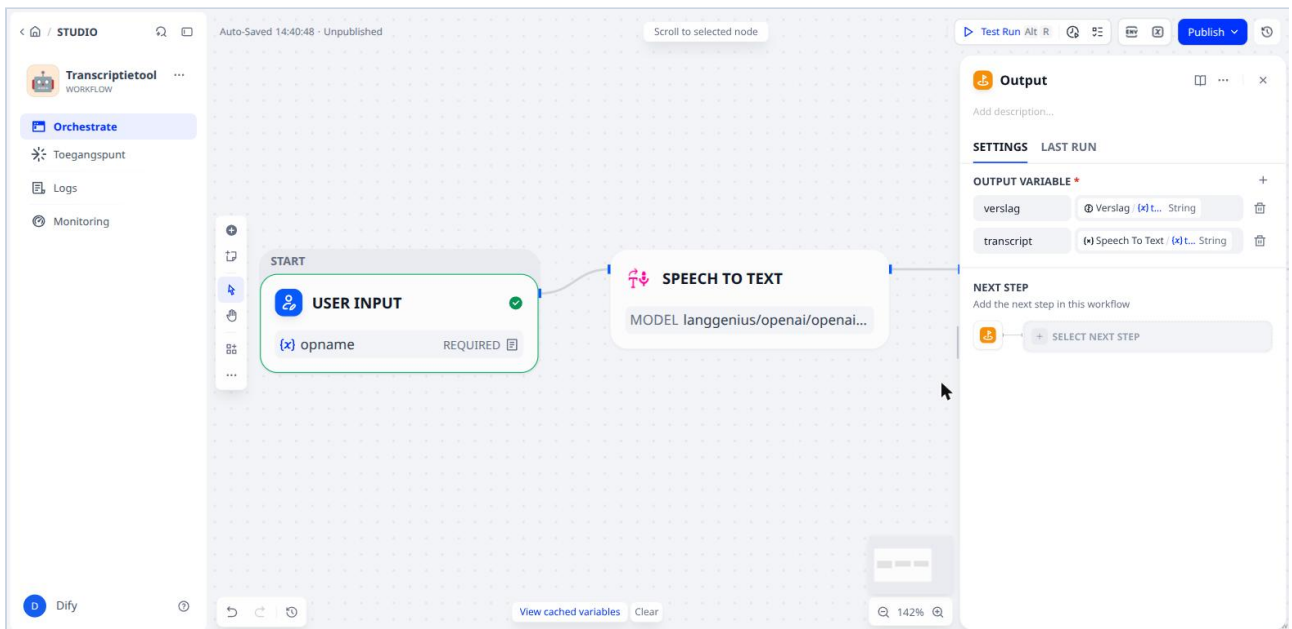
De lege Output-node

2. Klik op het plusje bij **OUTPUT VARIABLE** en maak twee variabelen.

Variable name (intypen)	Bron (kiezen)
verslag	Verslag / text
transcript	Speech To Text / text

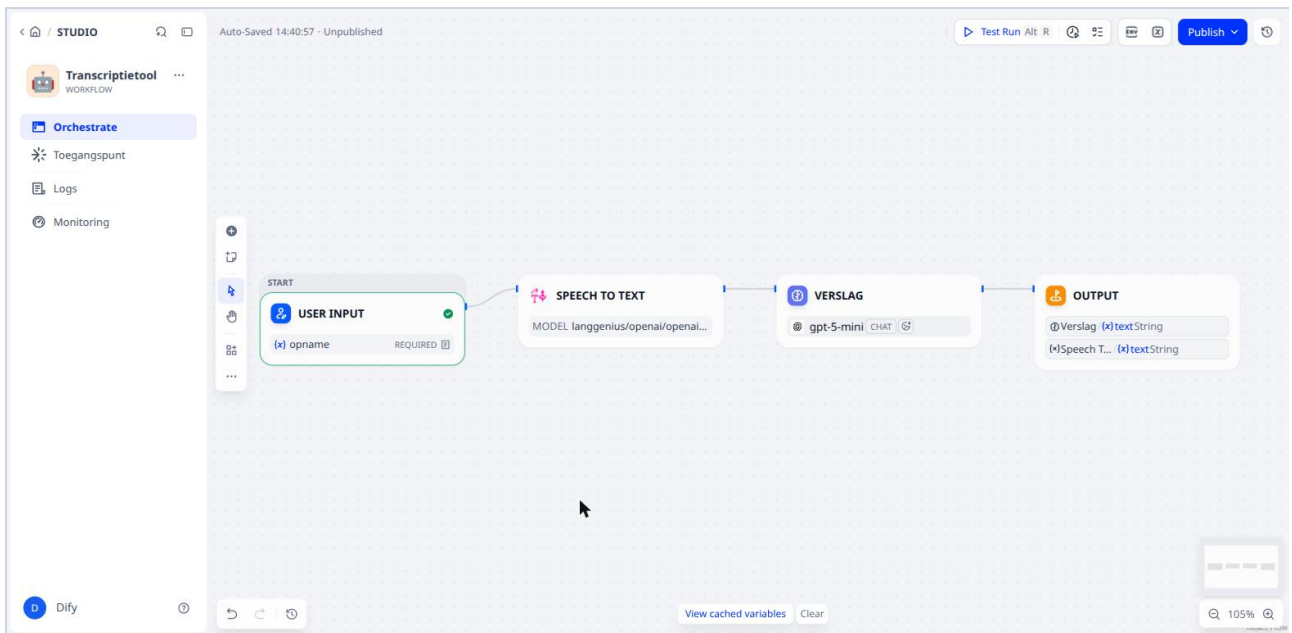


De bronnenlijst



Beide uitvoervariabelen

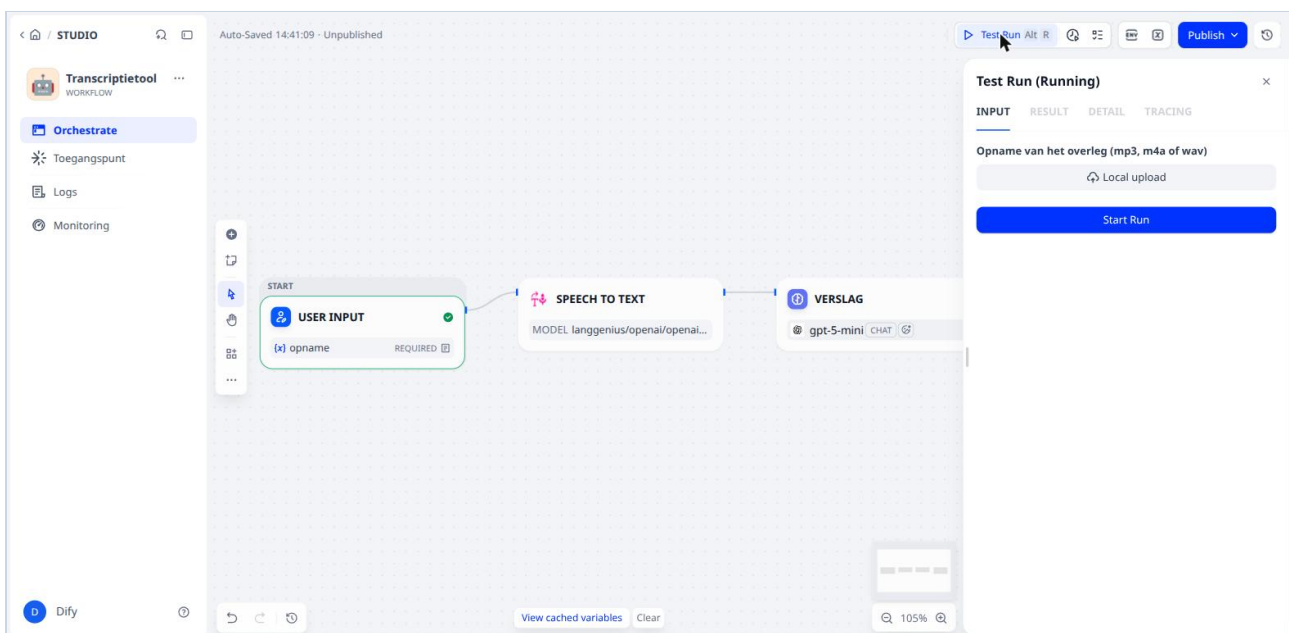
3. Sluit het paneel en druk op Ctrl+1. Je ziet de hele workflow: vier nodes.



User Input, Speech To Text, Verslag, Output

Stap 6. Test met de opname

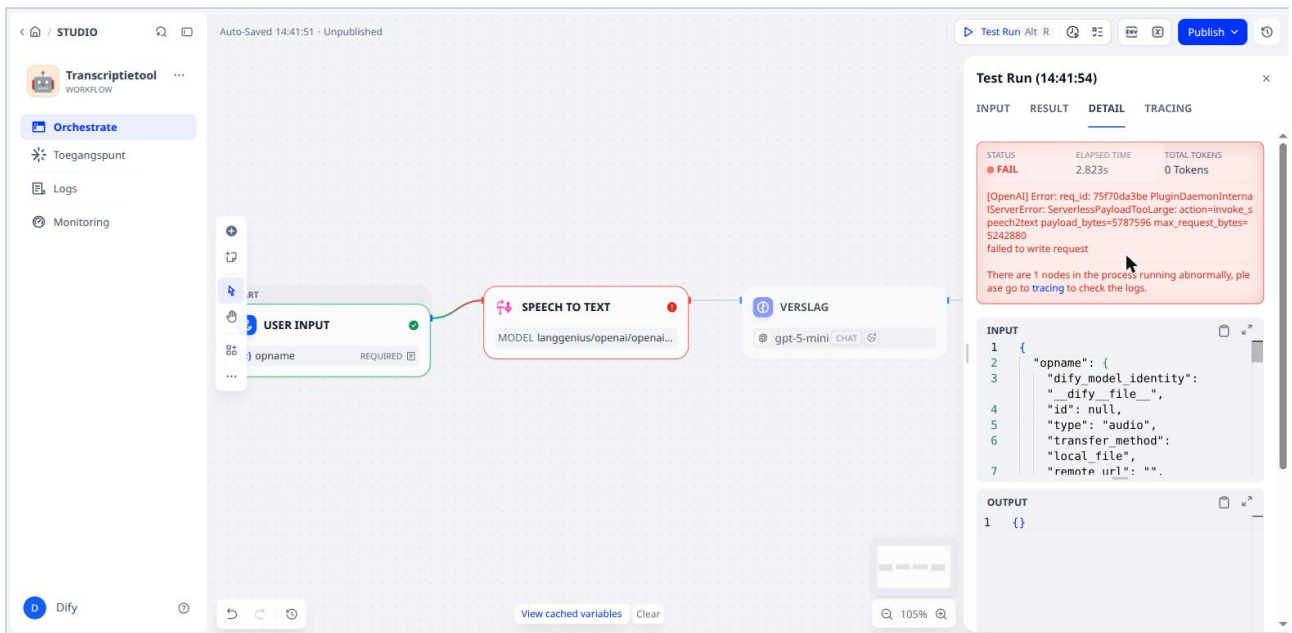
1. Klik rechtsboven op **Test Run**.



Het testpaneel met de uploadknop

2. Klik op **Local upload** en kies `sectieoverleg-opname-klein.mp3`.

Gebruik het bestand dat "klein" in de naam heeft. Het andere bestand in de map, `sectieoverleg-opname.mp3`, is dezelfde opname op 128 kbps. Die is 2,8 MB en dat is te groot: de tool stuurt audio als tekst (base64) naar de aanbieder, wat het bestand twee keer zo groot maakt, en de grens ligt op 5 MB. Je krijgt dan de foutmelding hieronder.

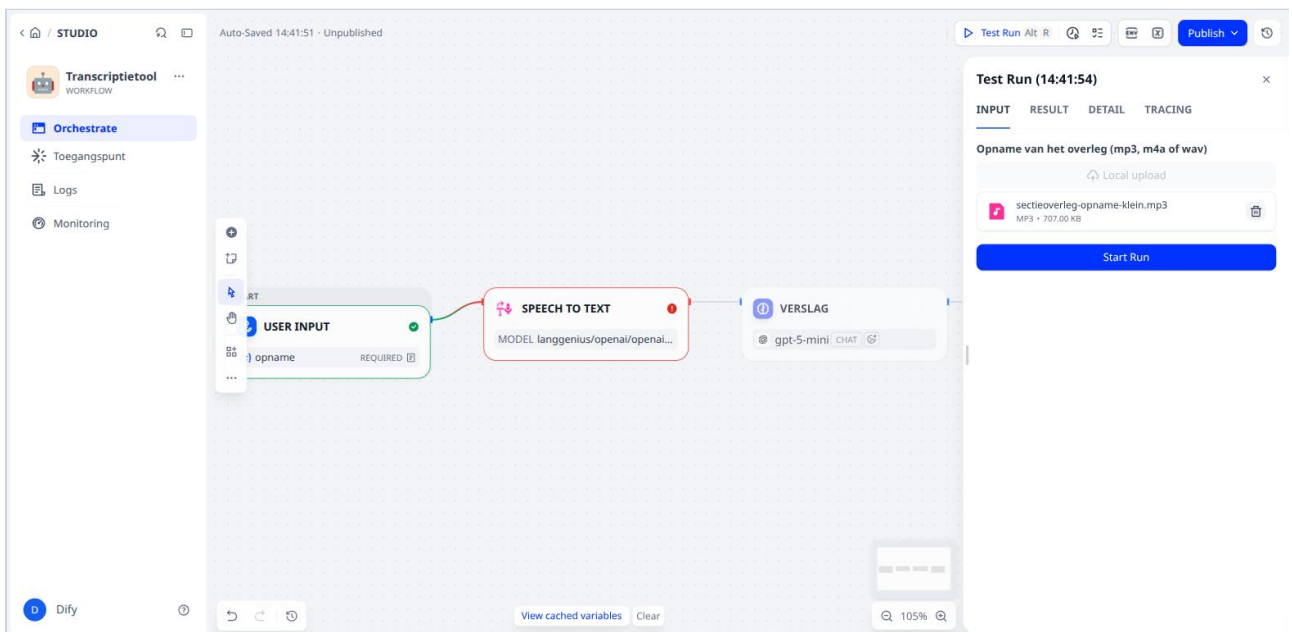


Te groot: ServerlessPayloadTooLarge

Vuistregel: **3 MB per opname**. Een half uur overleg past daar alleen in als je de opname eerst verkleint: mono, 16 kHz, 32 kbps geeft ongeveer 250 kB per minuut. Dat kan met elk audioprogramma, of met ffmpeg:

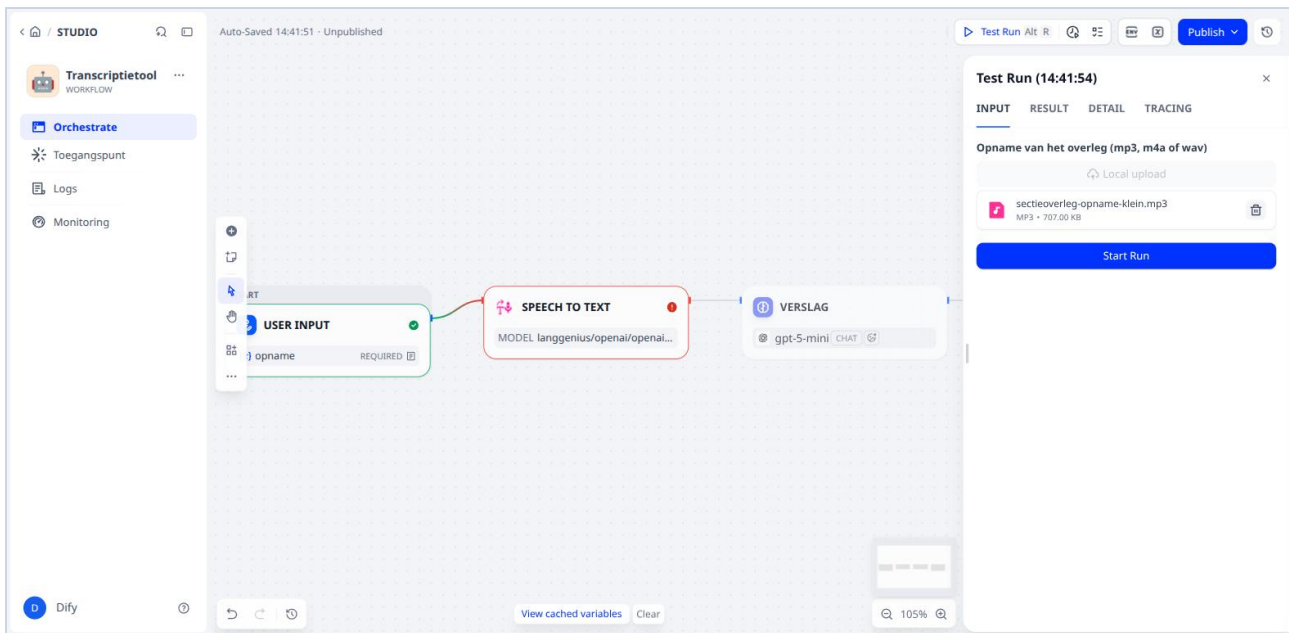
INTYPEN

```
ffmpeg -i opname.m4a -ac 1 -ar 16000 -b:a 32k opname-klein.mp3
```

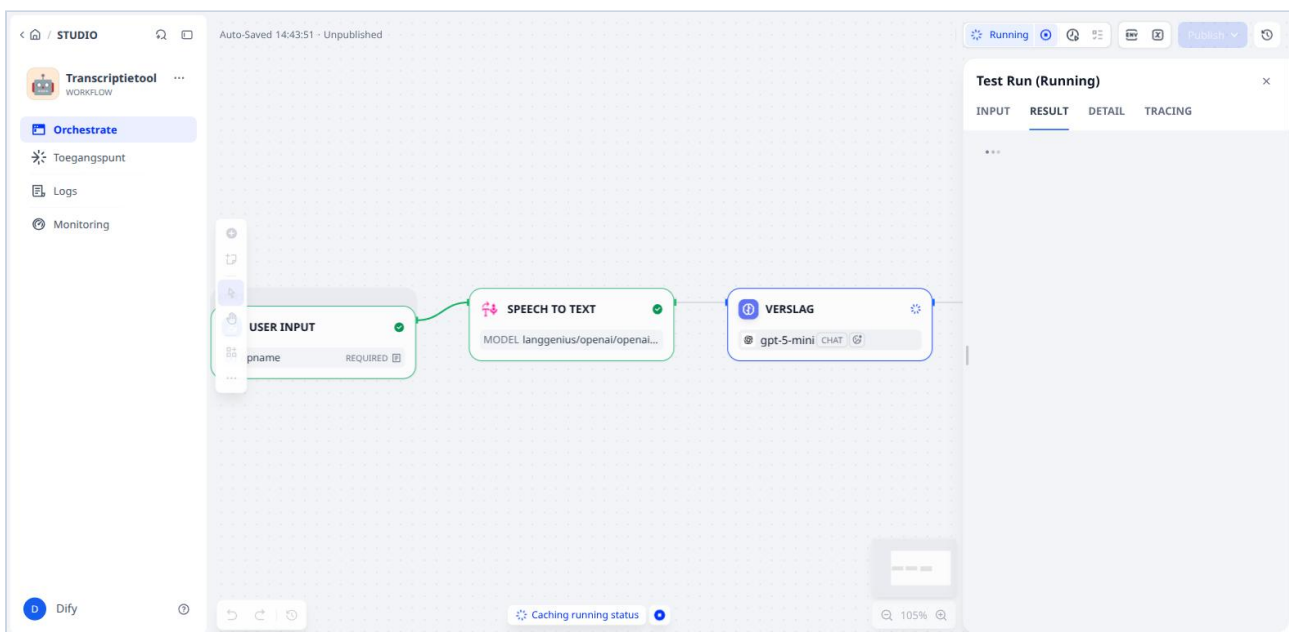


Het kleine bestand staat klaar

3. Klik op **Start Run**. De transcriptie van drie minuten audio duurt ongeveer een minuut.

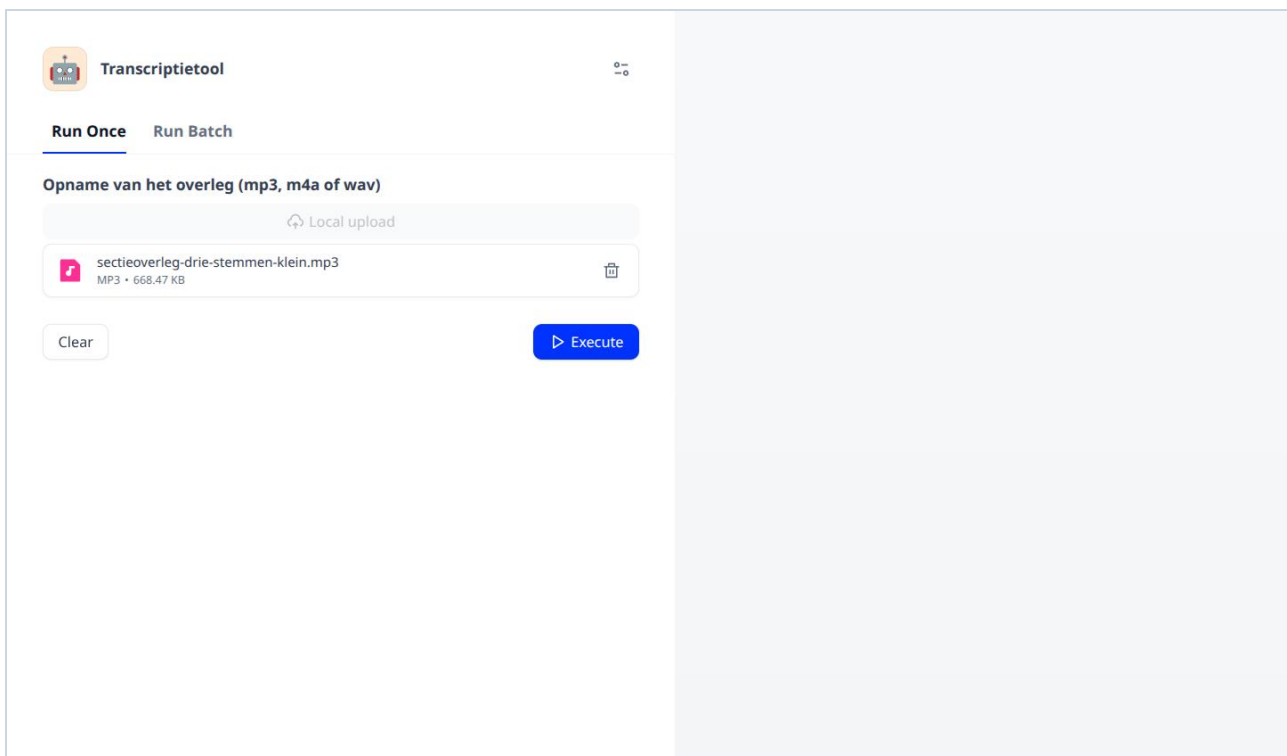


Speech To Text draait



De Verslag-node schrijft

4. Lees het resultaat onder **RESULT**.



De opname met drie stemmen staat klaar

- Kijk na afloop in **Logs** naar de run, tabblad **TRACING**, en klap **SPEECH TO TEXT** open. Onder OUTPUT staat wat de tool teruggeeft: één veld `text`, met een geregleinde op elke plek waar een andere stem begint.

The screenshot displays the Dify Studio interface. On the left, a sidebar contains navigation options: 'Orchestrator', 'Toegangspunt', 'Logs' (selected), and 'Monitoring'. The main area is titled 'Workflow Logs' and shows a table of workflow runs. The table has columns for 'START TIME', 'STATUS', 'RUN TIME', 'TOKENS', and 'END USER OR ACCOUNT'. Two runs are listed, both with a status of 'SUCCESS'. The right panel shows the 'Log Detail' for the selected run, with the 'TRACING' tab active. It displays the 'INPUT' and 'OUTPUT' for the 'SPEECH TO TEXT' step. The 'OUTPUT' section shows a JSON object with a 'text' field containing the transcribed text. The text is formatted with line breaks to indicate speaker changes.

START TIME	STATUS	RUN TIME	TOKENS	END USER OR ACCOUNT
09/20/2026 09:55:51 PM	SUCCESS	94.939s	4407	7f1b0c00-63dd-45cf-9caf-dea805edbfcs
09/20/2026 02:51:32 PM	SUCCESS	88.271s	4463	7f1b0c00-63dd-45cf-9caf-dea805edbfcs

```
{  "audio_file": {    "diffy_model_identity": "_diffy_file_",    "id": null,    "type": "audio",    "transfer_method": "local_file",    "remote_url": "",    "reference": "diffy-file-ref:ev3vZwNycnRfaw0101JUmIwH1jKZCOWMz31LTRh"  },  "process_data": {    "process": {}  },  "output": {    "text": "Sexy overleg Nederlandsch Dinsdag twee en twintig September. Aanwezig zijn Karin, Bram en Fatima. Karin zit voor. Goie middag allemaal, we hebben drie punten: de toetsweek van oktober, de nieuwe leesmethode voor de brugklas en de rolverdeling bij de open dag. Laten we beginnen met de toetsweek. Bram, jij had de concept planning gemaakt. Klopt.\nDe toetsleesvaardigheid voor de tweede klasse staat nu op maandag twaalf oktober."  }  }
```

De uitvoer van Speech To Text: tekst met gereleinden per spreker, geen namen

Dat is het eerlijke antwoord: het model **knip**t de tekst op sprekerwisselingen, maar de Speech To Text-tool van Dify geeft **geen labels** (Spreker 1, Spreker 2) door. Wie wat zei, leidt

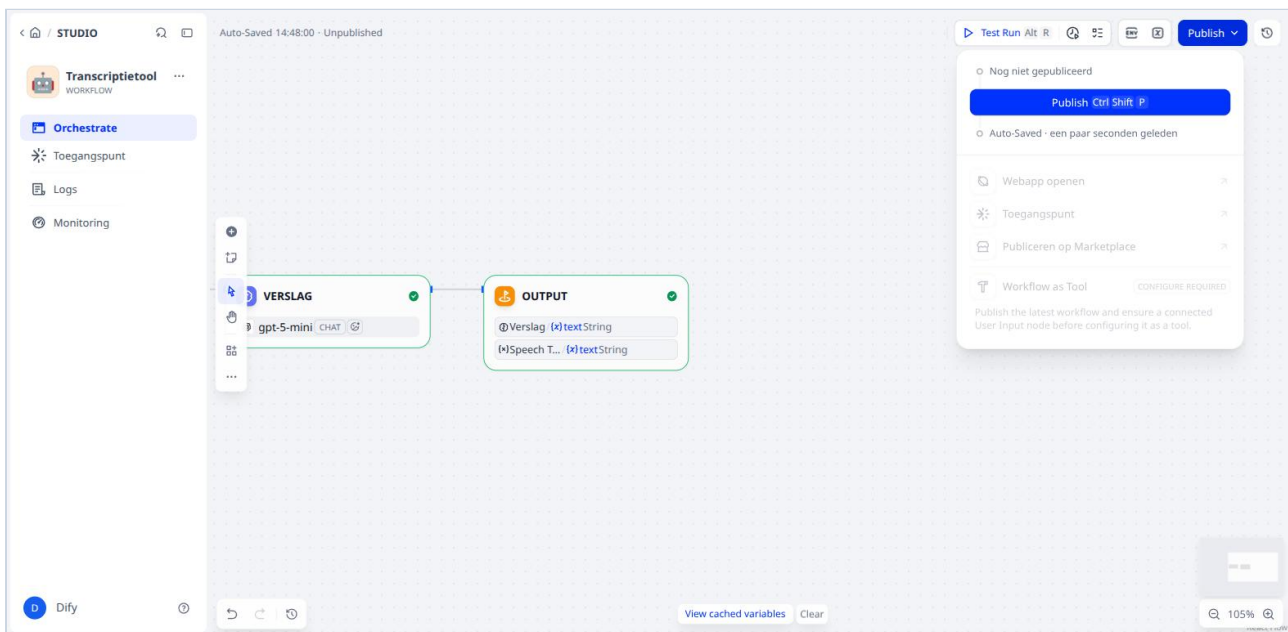
de Verslag-node af uit de inhoud, en dat lukt hier goed omdat de opening de namen noemt ("Aanwezig zijn Karin, Bram en Fatima") en omdat mensen elkaar in een overleg aanspreken. In een overleg zonder zo'n opening worden de actiepunten "niet genoemd".

Twee dingen vielen verder op. De transcriptie van de drie stemmen bevat meer fouten dan die van de ene stem ("Nederlandsch", een verhaspelde naam): de computerstemmen zijn getraind op Engels en spreken Nederlands met een accent, en dat hoort het spraakmodel. Echte Nederlandse sprekers gaan beter. En de doorlooptijd was hetzelfde, ruim een minuut voor drie minuten audio.

Zelf kiezen: echte sprekerlabels. Wil je "Karin:" en "Bram:" in het transcript, dan zijn er twee routes. De makkelijke: zet in de prompt van de Verslag-node "De regels in het transcript zijn sprekerwisselingen. Zet vóór elke regel de naam van de spreker als die uit de inhoud blijkt, anders Spreker 1, 2, 3." De precieze: vervang de Speech To Text-tool door een **HTTP Request**-node naar de spraak-API van OpenAI met `response_format=diarized_json`; die geeft per zin een spreker nummer en een tijd. Dat vraagt een sleutel in de node (via een omgevingsvariabele, nooit in de tekst) en is een volgende handleiding.

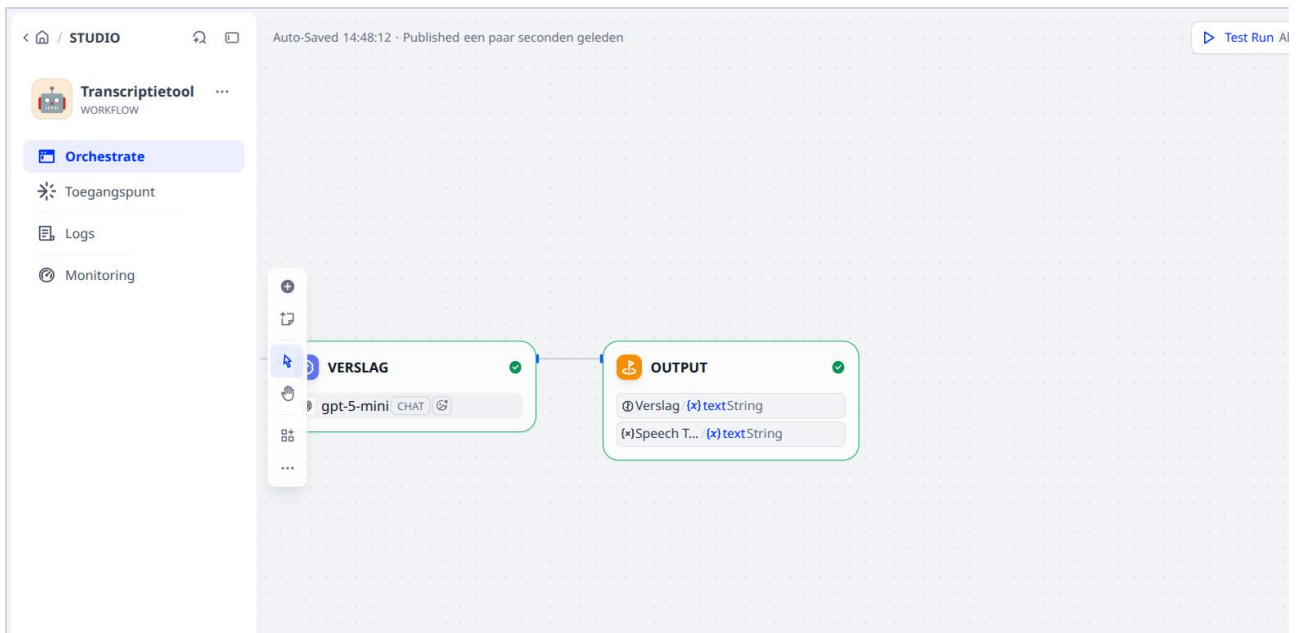
Stap 7. Publiceer

1. Klik rechtsboven op **Publish** en dan op **Publish**.



Het Publish-menu

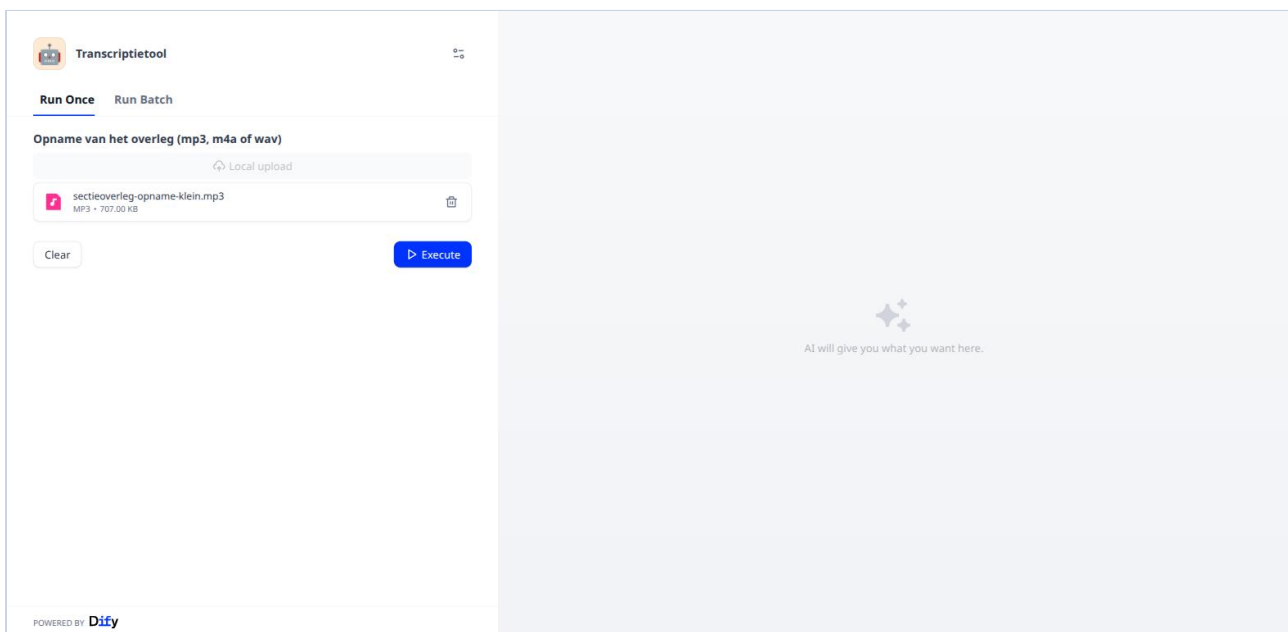
2. Open het menu opnieuw en klik op **Webapp openen**.



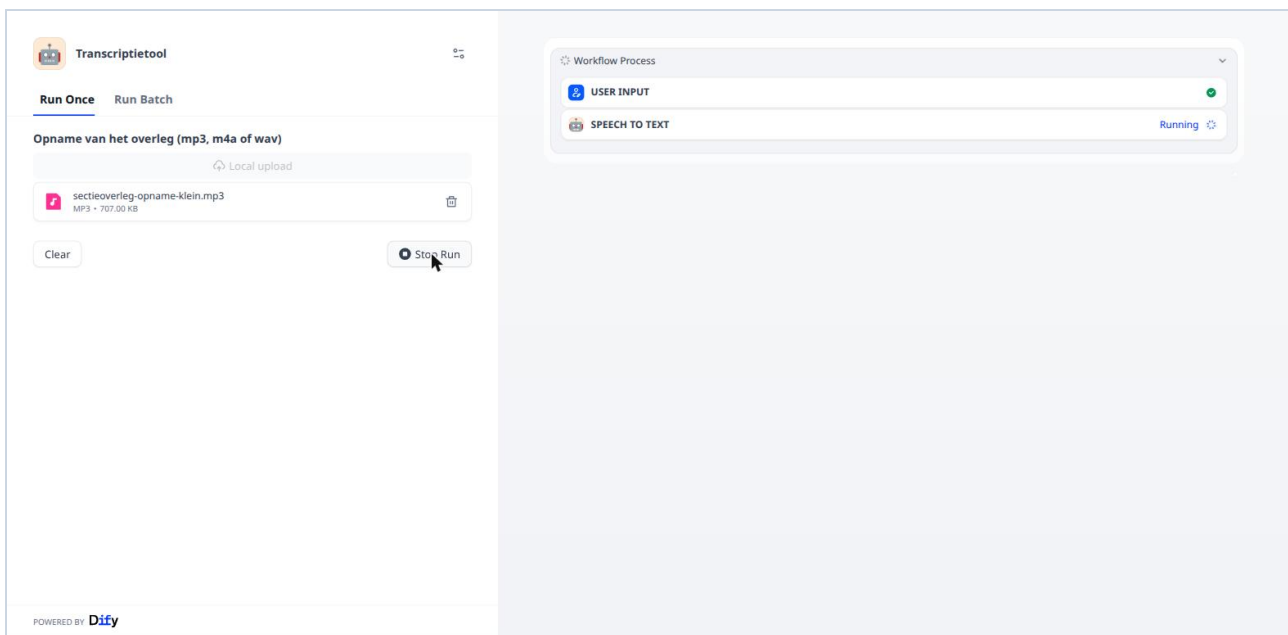
De webapp

3. Klik op **Local upload**, kies de kleine opname en klik op **Execute**.

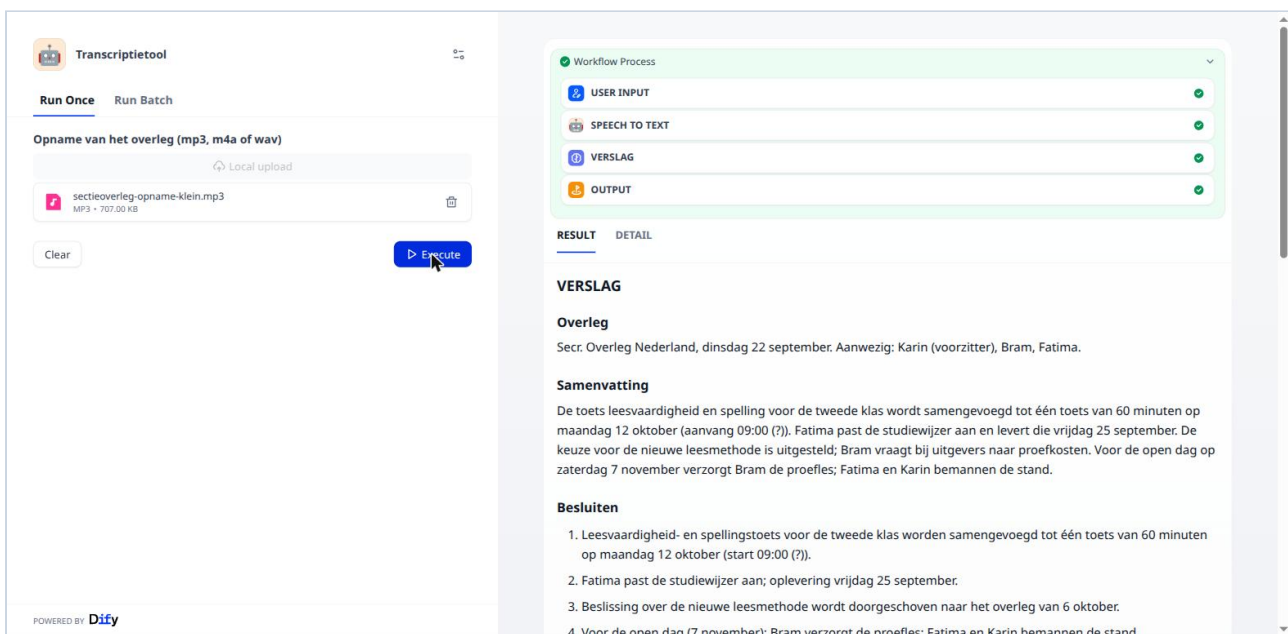
Na het uploaden schuift de knop Execute een stukje naar beneden. Kijk even waar hij staat voordat je klikt.



Bestand geüpload, Execute staat lager



De vier stappen lopen



Het verslag in de webbapp

Transcriptietool

Run OnceRun Batch

Opname van het overleg (mp3, m4a of wav)

Local upload

sectieoverleg-opname-klein.mp3
MP3 • 707.00 KB

Clear

Execute

POWERED BY Dify

Actiepunten

Wat	Wie	Wanneer
Studiewijzer aanpassen	Fatima	vrijdag 25 september
Proefkosten leesmethode opvragen bij uitgevers	Bram	deze week
Proeflesopzet sturen naar Karin	Bram	niet genoemd
Proefles verzorgen (open dag)	Bram	zaterdag 7 november
Stand Nederlands bemannen (open dag)	Fatima, Karin	zaterdag 7 november

Doorgeschoven

Keuze voor de nieuwe leesmethode — doorgeschoven naar overleg 6 oktober; reden: aanvullende informatie nodig over proefkosten en onzekerheid over tijd/organisatie (toetsweek).

Secr. Overleg Nederland, dinsdag twee en twintig september. Aanwezig zijn Karin, Bram en Fatima. Karin zit voor. Goedemiddag allemaal. We hebben drie punten: de toetsweek van oktober, de nieuwe leesmethode voor de brugklas, en de rolverdeling bij de open dag. Laten we beginnen met de toetsweek. Bram, jij had de concept planning gemaakt.

Klopt! De toets leesvaardigheid voor de tweede klasse staat nu op maandag twaalf oktober om negen uur zestig minuten. Ik stel voor, dat we die zoo laten. Wat ik wel wil voorstellen is, dat we de spellingstoets samenvoegen met de leestoets, dus één toets van zestig minuten in plaats van twee losse toetsen. Daar ben ik het mee eens. Twee losse toetsen in één week is voor de tweede klassers te veel.

Maar dan moet de studie wijze[r] wel worden aangepast, want daar staan nu nog twee toetsen in. Goed, dan besluiten we dat eengemaakte toets leesvaardigheid en spelling op maandag twaalf oktober, zestig minuten.

Fatima, pas jij de studie wijze[r] aan? Doe ik. Vrijdag heb ik hem klaar. Prima. Vrijdag vijf en twintig september dus. Dan punt twee, de leesmethode. We hebben twee proefpakketten gekregen. Bram heeft ze bekeken.

Ja, en eerlijk gezegd vind ik het lastig kiezen. Het ene pakket heeft veel betere teksten voor zwakke lezers, het andere heeft een digitale omgeving die veel beter werkt. Ik zou eigenlijk willen dat we ze allebei een periode uitproberen in één klas. Kunnen we dat wel betalen, en hebben we daar de tijd voor met de toetsweek eraan? Ik denk, dat we dit niet vandaag moeten beslissen. Ik stel voor, dat we dit doorschuiven naar het volgende overleg

De actiepunentabel, met kopieer- en downloadknop

Rechtsboven bij de tabel staan een kopieerknop en een downloadknop. Daarmee zet een collega de actiepunten zo in een mail of een spreadsheet.

Het voorbeeld uit deze handleiding: <https://udify.app/workflow/7N5dNg8fD9qXHA80>.

Controleren of het gelukt is

- De workflow heeft vier nodes: User Input, Speech To Text, Verslag, Output
- Het invoerveld accepteert alleen audio (geen Image aangevinkt)
- Test Run met de kleine opname levert een verslag met vijf kopjes en daaronder het transcript
- In de webapp komt hetzelfde resultaat, met VERSLAG als kop
- Onzekere namen staan met een vraagteken tussen haakjes

Als het misgaat

Wat je ziet	Wat er aan de hand is
<code>ServerlessPayloadTooLarge ... max_request_bytes=5242880</code>	Het audiobestand is te groot. Verklein naar 32 kbps mono (stap 6)
Speech To Text: geen model in de lijst	Geen eigen OpenAI-sleutel, of Gebruiksprioriteit staat op AI-tegoeden. Integraties → Model Provider → OpenAI → Config
De workflow loopt vast op Speech To Text bij een foto	Image staat nog aan bij het invoerveld (stap 2)
Het verslag noemt mensen die niet in het overleg zaten	De spraakherkenning heeft een naam verkeerd verstaan. Kijk in het transcript onderaan; vaak staat de juiste naam er wel bij met een vraagteken
Verslag en transcript lopen in elkaar over	De regel "Begin je antwoord met de regel # VERSLAG" ontbreekt in de instructie
"Upgrade to create more apps"	Sandbox-limiet van vijf apps. Verwijder een oude app via de drie puntjes op de kaart
Een mp3-link uit een testrun geeft <code>403 forbidden</code>	Downloadlinks van Dify-bestanden zijn vijf minuten geldig. Download meteen, of draai de test opnieuw
Het transcript heeft geen sprekeramen, alleen regeleinden	Zo werkt de Speech To Text-tool: diarize knipt, maar labelt niet in Dify. Zie Sprekers herkennen bij stap 6
Melding over temperature	Het model ondersteunt die parameter niet. Negeren

Maak hem van jou

De opname is verzonnen, de indeling van het verslag is onze keuze. Dit hoofdstuk helpt je de tool om te bouwen naar jullie overleggen, binnen het privacykader aan het begin van deze handleiding. Van makkelijk naar moeilijk.

Aanpassing	Waar	Moeilijkheid	Wat je ervoor terugkrijgt
Andere kopjes in het verslag	Stap 4, prompt	makkelijk	Een verslag in de vorm die jullie al gebruiken
Taal en toon (formeel, informeel, Engels)	Stap 4, prompt	makkelijk	Een verslag dat je zonder herschrijven kunt delen
Een ander spraakmodel (whisper-1 is goedkoper)	Stap 3, instellingen van de tool	makkelijk	Lagere kosten; sprekerwisselingen als geregeleinden verlies je dan
Sprekernamen in het transcript	Stap 6, prompt van Verslag, of een HTTP Request-node (zie het groene kader bij Sprekers herkennen)	gemiddeld	Een transcript met "Karin:" en "Bram:" ervoor
Een tweede uitvoer: kort bericht voor wie er niet was	Extra LLM-node na stap 4, extra uitvoervariabele in stap 5	gemiddeld	Twee producten uit één opname
Actiepunten als aparte lijst die je kunt kopiëren naar een takenlijst	Extra LLM-node met alleen de tabel	gemiddeld	Minder overtypen na het overleg
Een invoerveld "type overleg" dat de indeling kiest	Stap 2 plus een IF/ELSE-node of een prompt met voorwaarden	gemiddeld	Eén tool voor meerdere soorten overleg
Langere opnames dan 5 MB	Speech To Text vervangen door een HTTP Request-node naar de spraak-API	moeilijk	Een uur overleg in één keer; zie Beheer
Draaien in de SOML-omgeving met Vertex AI	Export DSL, importeren, modellen opnieuw koppelen	moeilijk	De route die voor echte opnames verplicht is

Drie opdrachten

- 1. Jullie indeling.** Pak het laatste verslag dat iemand in jouw team met de hand schreef. Zet de kopjes daarvan in de prompt van stap 4, met per kopje één zin uitleg. Draai de fictieve opname opnieuw en vergelijk met het handgeschreven verslag: wat mist, wat is te veel?
- 2. Maak een eigen testopname.** Schrijf met je team een fictief script van tien minuten over een echt onderwerp (zonder echte namen), en maak er via Bijlage A een opname van. Zo test je de tool op jullie jargon zonder dat er een echte stem in Dify komt.

3. **Bereid de overstap voor.** Exporteer de workflow (Export DSL) en schrijf op een half A4 wat er nodig is om hem in de Experimenteerruimte te draaien: welk model, welke sleutel, wie de eigenaar is, hoe lang audio bewaard blijft. Dat half A4 is precies wat de use case-beoordeling van je vraagt.

Checklist voordat je eigen materiaal gebruikt

- [] Er gaat geen echte opname in Dify-cloud; alleen fictieve of synthetische audio
 - [] Voor echte opnames: de tool draait in de SOML-omgeving via het centrale project, en de deelnemers hebben ingestemd
 - [] Audio wordt verwijderd na de transcriptie en er is vastgelegd wie dat controleert
 - [] Het scenario met leerlingen (leerlingbesprekingen, gesprekken met leerlingen) wacht op de DPIA
 - [] Iemand leest het verslag na voordat het rondgaat; de vraagtekens in het verslag zijn daarvoor bedoeld
 - [] Er is één eigenaar van de tool die de prompt bijhoudt en opnieuw publiceert
-

Beheer

Wat Dify bewaart. Elke upload en elk resultaat staat in **Logs** van de app, met het audiobestand erbij. Voor het prototype met de fictieve opname is dat geen probleem. Zodra er ooit een echte opname in zou gaan, is dit de reden dat Dify-cloud niet mag: je kunt de audio niet verwijderen zonder de hele log te wissen, en het bewaarbeleid van de use case (audio weg na transcriptie) is dan niet na te komen. Voor echt gebruik hoort de tool in de SOML-omgeving, met een bewaartermijn die is vastgelegd.

Kosten. Via je eigen sleutel: gpt-4o-transcribe-diarize kost ongeveer 0,6 cent per minuut audio, whisper-1 de helft. Een overleg van 45 minuten komt op 20 tot 30 cent inclusief het verslag. Dify-credits worden niet aangesproken zolang OpenAI op API-sleutel staat.

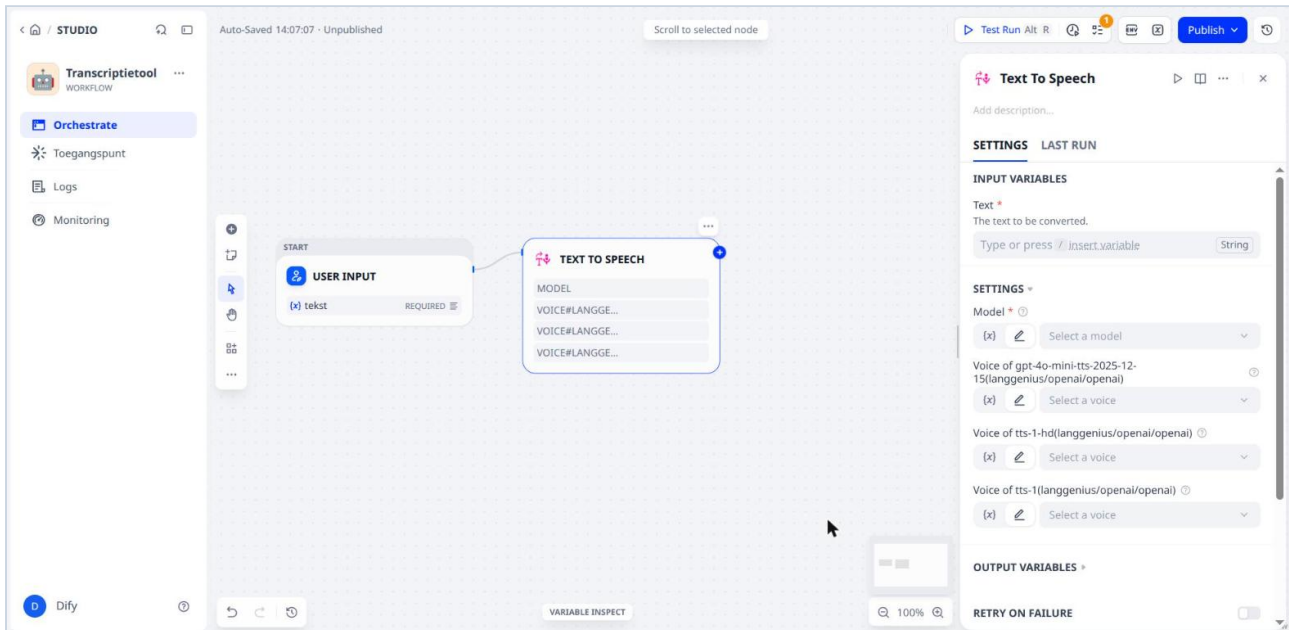
Langere opnames. De 5 MB-grens van de tool is hard. Een uur overleg op 32 kbps is 14 MB en past dus niet. Twee routes: de opname in stukken van een kwartier knippen en apart draaien, of de Speech To Text-tool vervangen door een HTTP Request-node die rechtstreeks naar de spraak-API van de aanbieder gaat (grens 25 MB). Dat laatste is een volgende handleiding.

Van prototype naar de Experimenteerruimte. De workflow is via **Export DSL** (drie puntjes op de kaart in Studio) als bestand op te slaan en in een ander Dify-account te importeren. De modellen moeten daar opnieuw gekoppeld worden aan de sleutels van dat account. Voor de Gemini-route: Model Provider → Gemini → Add API Key met de sleutel uit het centrale SOML-project.

Bijlage A. Een testopname maken zonder echte stemmen

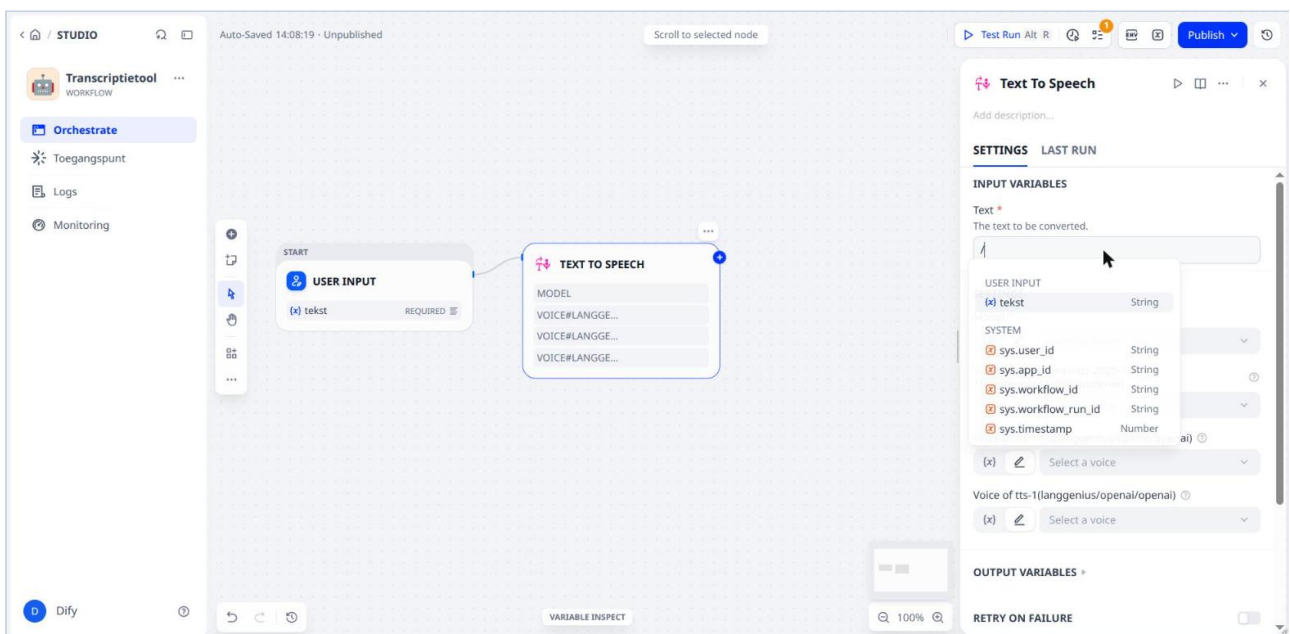
Voor een prototype heb je een opname nodig zonder persoonsgegevens. De makkelijkste route is een fictief script laten voorlezen door dezelfde Audio-tool, maar dan andersom: **Text To Speech**. Zo is de opname bij deze handleiding gemaakt.

1. Maak een workflow met een **User Input** met één veld: type **Paragraph**, variabele `tekst`.
2. Kies als volgende stap **Tools** → **Audio** → **Text To Speech**.



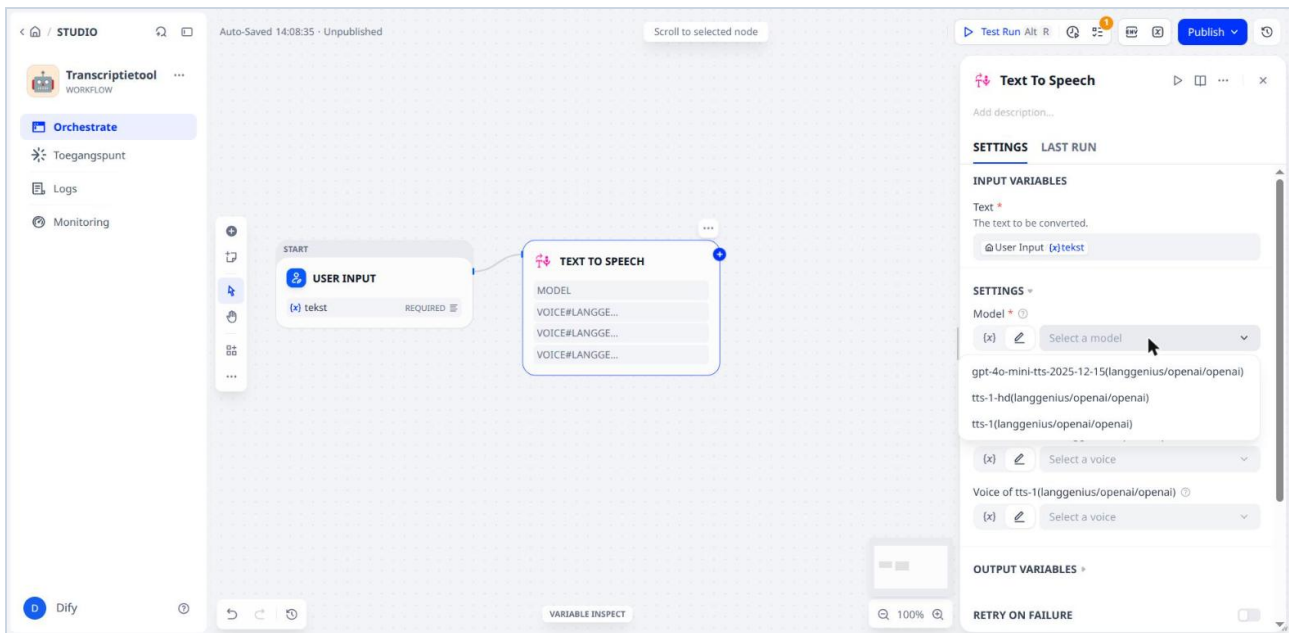
De Text To Speech-node

3. Klik bij **Text** in het veld, typ `/` en kies **tekst**.



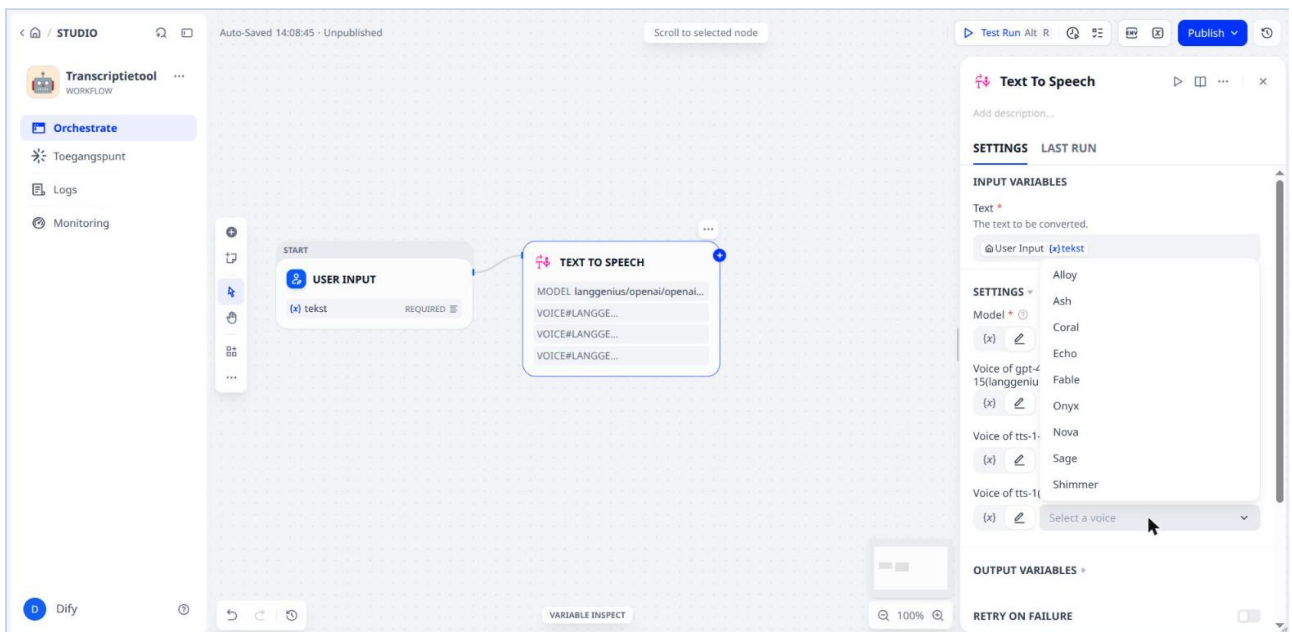
De variabele kiezen

4. Kies bij **Model** een spraakmodel; **tts-1** is het goedkoopst.



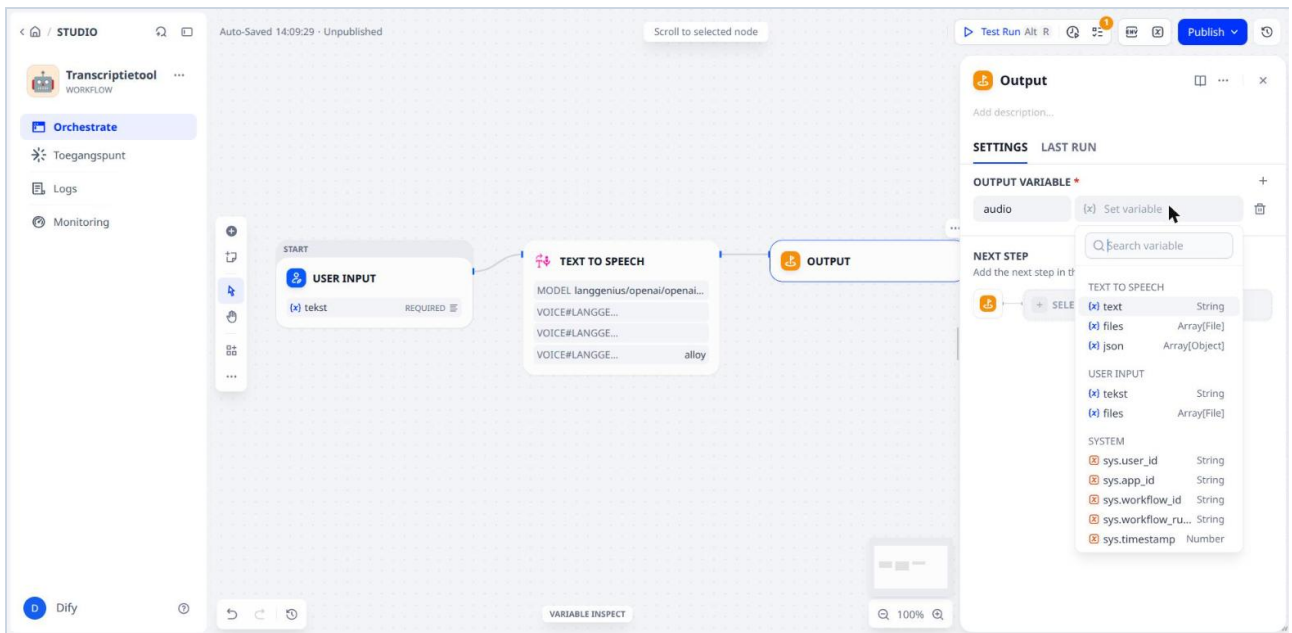
De spraakmodellen

5. Kies bij **Voice of tts-1** een stem, bijvoorbeeld **Alloy**.



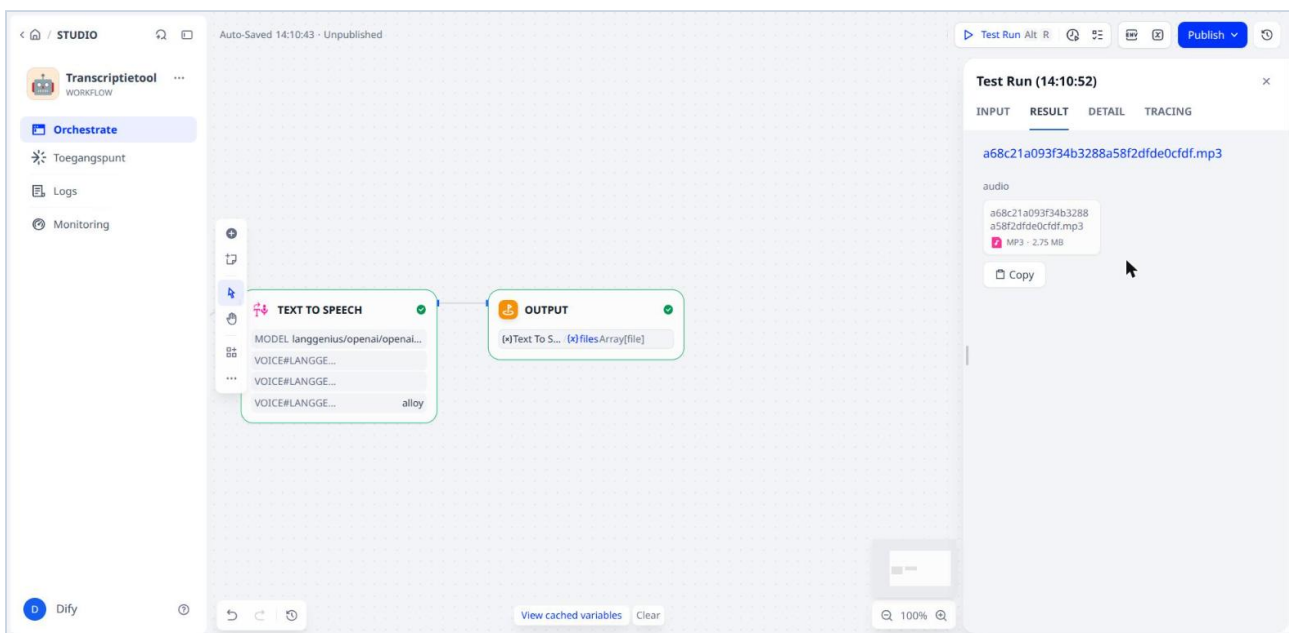
De stemmen

6. Voeg een **Output**-node toe met variabele **audio** → Text To Speech / **files**.



De Output-node met het audiobestand

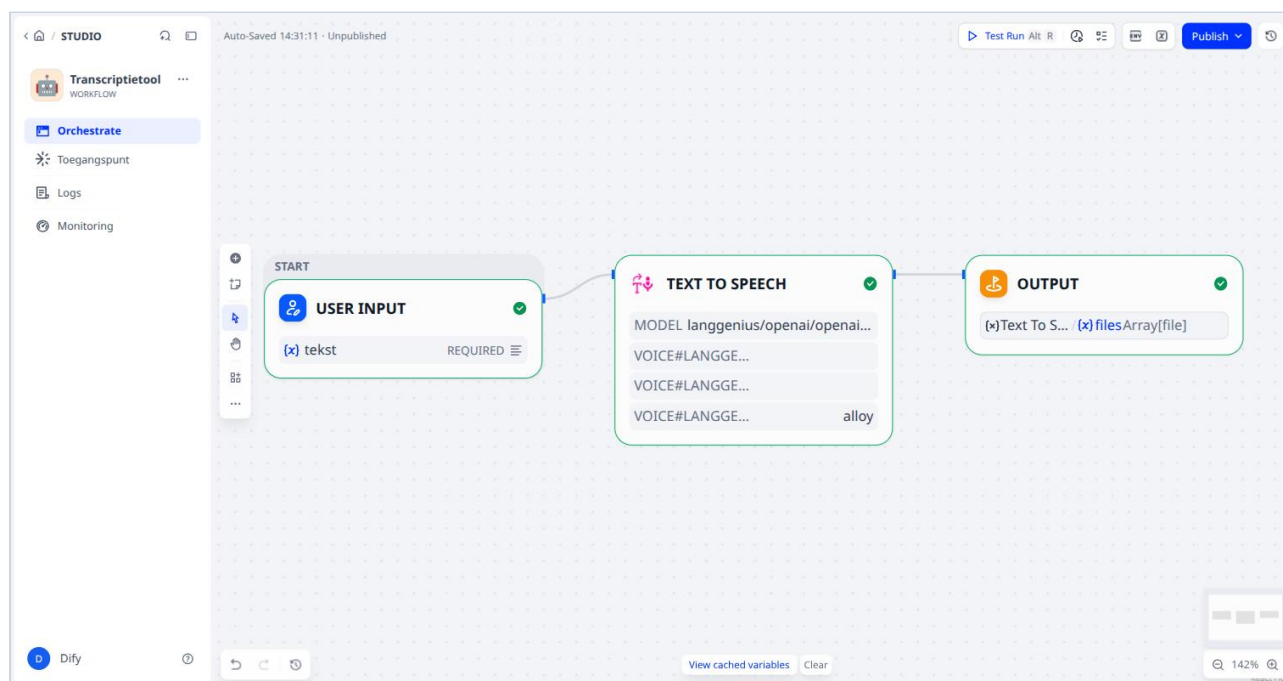
7. Klik op **Test Run**, plak het script (sectieoverleg-script.txt) en klik op **Start Run**. Na een halve minuut staat er een mp3 in het resultaat.



De opname is klaar

8. Klik meteen op de bestandsnaam en sla het bestand op. De link werkt vijf minuten.

Kosten: ongeveer 4 cent voor drie minuten spraak. Het resultaat is één stem die alle sprekers leest; voor een test van transcriptie en verslag is dat genoeg. Een sandbox-account heeft plek voor vijf apps, dus verwijder deze hulpworkflow daarna, of bouw hem om tot de Transcriptietool zoals in deze handleiding is gedaan.



De hulpworkflow: User Input, Text To Speech, Output

Licentie: CC BY-SA 4.0 (<https://creativecommons.org/licenses/by-sa/4.0/deed.nl>). Je mag dit document kopiëren, bewerken en hergebruiken, ook op je eigen school, als je de bron noemt (Guido van Dijk, LeX Consultancy B.V. en AI Netwerk Limburg, <https://git.lexconsultancy.nl/guido/dify-handleidingen>) en je bewerking onder dezelfde licentie deelt. Het logo van AI Netwerk Limburg valt buiten de licentie; namen van producten zijn merken van hun eigenaren. Scholen, personen en documenten in de voorbeelden zijn verzonnen.

Eigenaar en contact: Guido van Dijk, LeX Consultancy B.V. en AI Netwerk Limburg · guido@l-e-x.nl · guido@ai-netwerk-limburg.nl



AI Netwerk Limburg · LeX Consultancy B.V. · Bouwen in Dify: van opname naar verslag