



The second European Language Resource Coordination (ELRC) workshop in Luxembourg



European
Resource Co
Connecting E

The potential of Language Technology and AI

ELRC Workshop Luxembourg

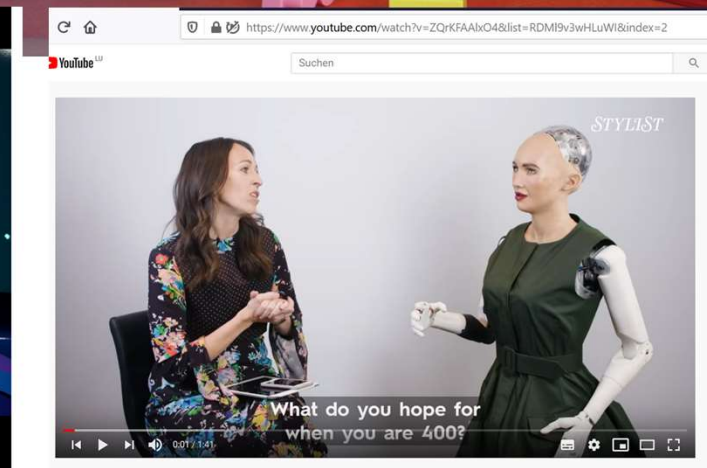
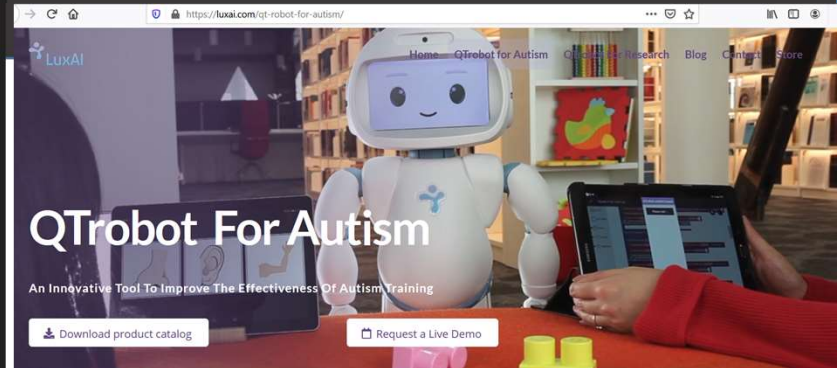
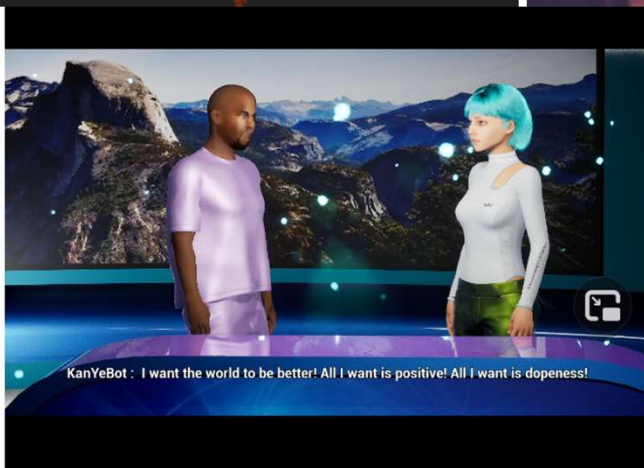
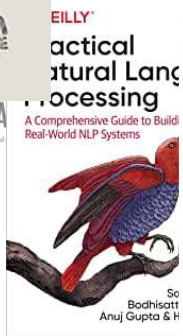
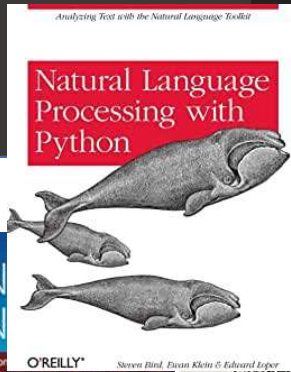
11 December 2020

Christoph Schommer

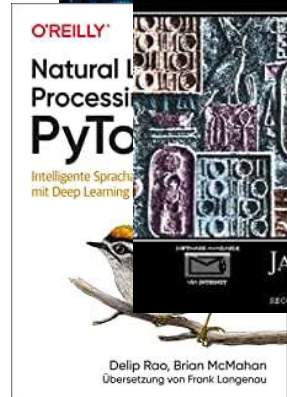
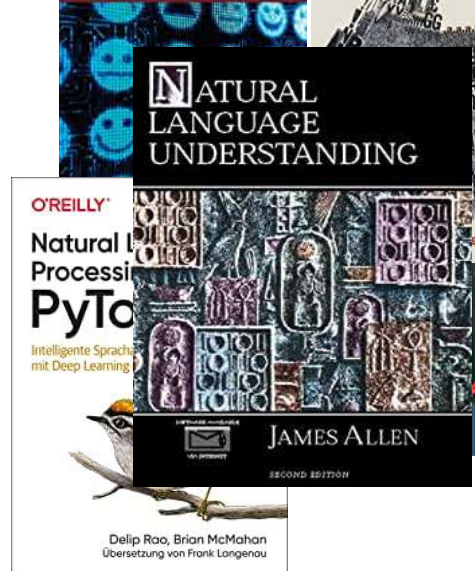
Dept. of Computer Science

University of Luxembourg

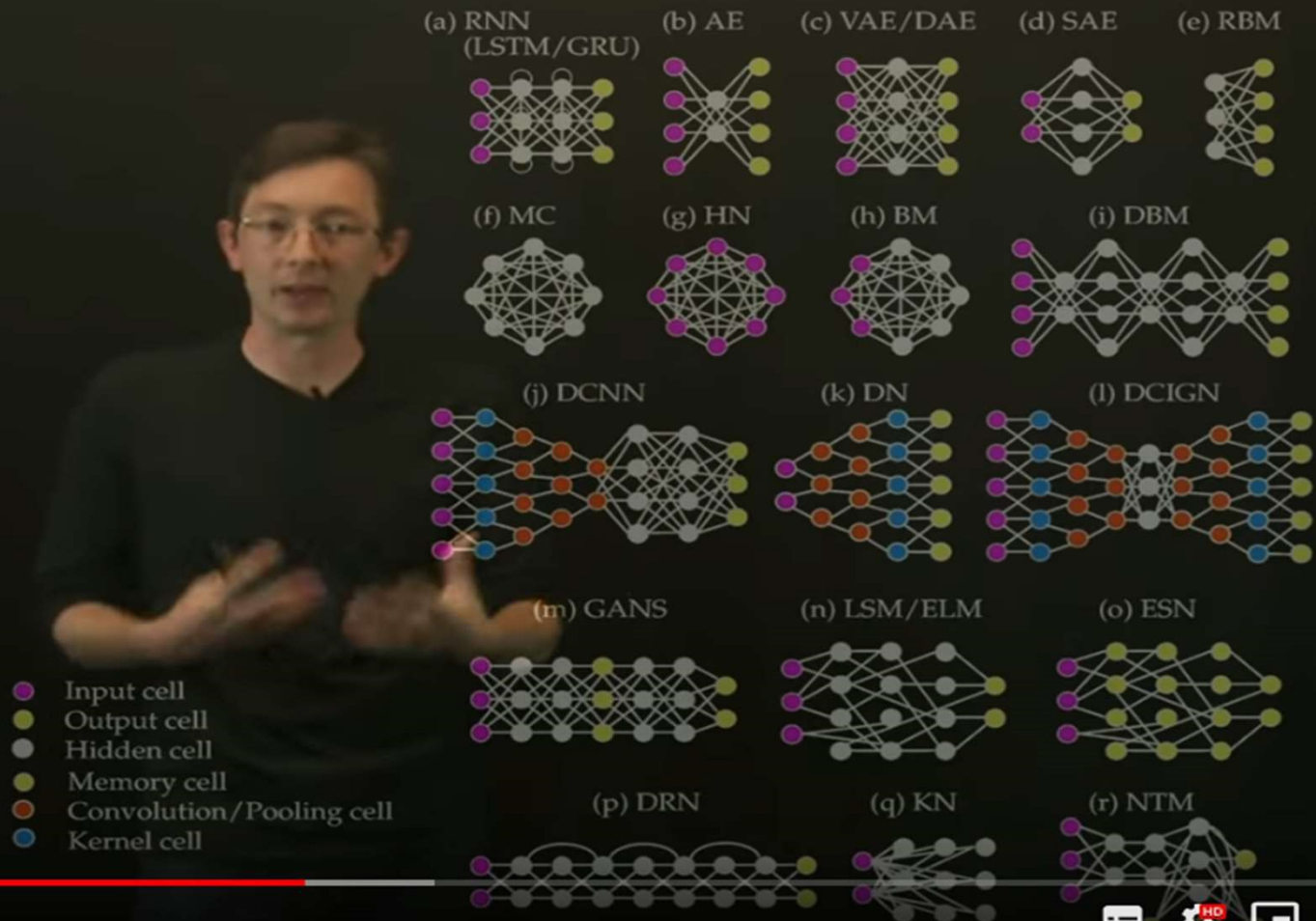
Natural Language Processing



Sophia the robot interview: Sophia the robot answers Stylist's philosophical questions



Deep Learning for AI ?!



Can AI write poetry?

- <https://www.youtube.com/embed/WGt8MkeGpNA> 4min 55 sec
- <https://www.newyorker.com/culture/annals-of-inquiry/the-mechanical-muse>

“Good night! Tomorrow is another day :)”

*I wanna be a little kid again.
I'm feeling kinda empty on the low.
You should unwind a little now and then.
Team Stacie looking like a sleepy hoe.*

*Back to the Sunshine State. The devil is
a lie. I hate myself a lot sometimes,
I mean, possessive, holy shit, this is
the second time. I'm always catching dimes.*

*I'm not the only one, I'm pinning this
again. I love a windy sunny day.
Not coming out until tonight. I miss
the happy me. I gotta find a way.*

DeepHouse

http://10.240.2.56/

David Matos FREITAS
Fabio DI BIASE
Gabriela Vieira PINTO
Gilles CHEN
Joshgun Sirajzade
Christoph Schommer

Covid-19 Warehouse

Dashboard Covid-cases USA States EU Countries Tweets

24 / 07 / 2020

USA STATES EU Countries

- USA STATES
- American Samoa
- Guam
- Northern Mariana Islands
- Puerto Rico
- Virgin Islands
- Alabama
- Alaska
- Arizona
- Arkansas
- California
- Colorado
- Connecticut
- Delaware
- District of Columbia
- Florida
- Georgia
- Hawaii
- Idaho
- Illinois

USA

State:	New Mexico
Date:	2020-07-26

Europe

State:	Portugal
Date:	2020-07-26

|| unlu.wbex.com is sharing your screen. Stop sharing Hide

Tweets

Kirsty Duncan 36137
from Etobicoke North
Thinking of our wonderful community of #EtobicokeNorth tonight and always. We've come so far together during...
<https://t.co/my9uMFKRg5>

Rocket Lawyer 13928
from San Francisco
As a result of #COVID19, the federal government enacted a temporary eviction moratorium to provide relief for...
<https://t.co/P94W0Oydmf>

Mayor Marty Walsh 323916
from Boston, Mass.
I declared a heat emergency effective Sunday-Tuesday for the @CityofBoston. If you decide to leave your home, pleas...
<https://t.co/h1FDuckm1>

City of Miami Beach 107650
from Fun & Sun Capital of the World
City of Miami Beach Exte...
State of Emergency Decl...

<https://t.co/s3WtwEutAq>

New Straits Times 716397
from Malaysia
#NSTnation The Health Ministry reported 13 new #Covid19 cases in the country as of noon today.
<https://t.co/xlpaGuv2yt>

Hindustan Times 7649600
from India
President Kovind donates Rs 20 lakh to Army hospital to buy equipment to combat #Covid19
<https://t.co/yRaNGT1h5A>
<https://t.co/mR06158nMm>

WION 234126
from India
Take a look at India's daily caseload trajectory and the number of coronavirus fatalities. #Coronavirus #COVID19...
<https://t.co/zlFzJLC7P4>

HT Delhi 16823
from Delhi and NCR
At present, there are 11,904 active #Covid19 cases in Delhi
<https://t.co/gEEmeMkRYx>

NewsMobile 15334
from India



TWEETS

Username

Username of a tweeter to identify who wrote the tweet

Followers

The date of the tweet, to link it on our website

Text

The main part of the tweet

Hashtags

Hashtags used by the user



Kirsty Duncan 36137
from Etobicoke North
Thinking of our wonderful community of #EtobicokeNorth tonight and always. We've come so far together during...
<https://t.co/my9uMFKRg5>

STRIPS

<http://sentilux.uni.lux/stripsannotation/sentences.php>

Joshgun Sirajzade
Christoph Schommer
Christoph Purschke
Peter Gilles
Daniela Gierschek

The image shows a side-by-side comparison of two browser windows displaying the STRIPS web application. The left window is at the URL `sentilux.uni.lux/stripsannotation/firstfile.php` and has the title "Lëtzebuergesch Sätz bewäerten". It contains instructions in Luxembourgish on how to use the tool to evaluate sentences. The right window is at the URL `sentilux.uni.lux/stripsannotation/sentences.php` and has the title "Bewäert den ervirgehuewene Saz". It shows a sample sentence in Luxembourgish with a highlighted segment and three buttons for evaluation: "negativ", "neutral", and "positiv". Below these buttons is a text input field labeled "Saz iwwersprangen" and a blue button labeled "Uleedung". At the bottom of the right window, it shows the user's name as "Utilisateur: username" and a counter "Schonns 0 Sätz bewäert". A footer at the very bottom of the right window provides information about the project and the authors.

Lëtzebuergesch Sätz bewäerten

't wier flott, wann Dir eis bei der Analys vu Bewäertungen am Lëtzebuergeschen hëllef gëngt. Mat Ärer Hëllef probéiere mir, eng Software ze entwéckelen, déi automatesch decidéiert, ob e Saz positiv, negativ oder neutral ze verstoen ass.

Uleedung

No an no kritt Dir Sätz ugewise. Liest de Saz an decidéiert, ob e fir lech positiv, negativ oder neutral ass. Klickt dofir op den entspreichende Button. Soubal Dir op de Button geklickt hutt, gëtt Är Äntwert gespäichert an Dir kommt net méi bei dese Saz zeréck.

Duerno kritt Dir den nächste Saz gewise. Bewäert souvill Sätz, wéi Der wëllt oder kënnt. Dir kënnt zu all Moment ophalen a méi spéit rëm virufueren.

Wichtig: Huel bei der Bewäertung d'Perspektiv vun der Persoun, déi geschriwwen huet (Auteur), an.

Klickt op „iwwersprangen“:

- wann de Saz op Däitsch, Franséisch oder enger anerer Sprooch ass. (Et sollen nëmmen déi lëtzebuergesch Sätz annotéiert ginn.)
- wann de Saz onverständlech ass (z.B. *2008-12-02-15:09-MaxMustermann*).
- wann Dir lech net sécher sidd, ob de Saz positiv, negativ oder neutral ass.

START

Bewäert den ervirgehuewene Saz

Et gesät én mat der Versetzung vum Colonel, dass wann d'CSV e Beamten erlédigen well, dass dé Beamten absolut keng wirklech Rechter huet, fir sech géint dén "Beamtemord" ze wieren. **An esou Leichen gin et der sécher e puer honnert beim Staat, déi dem politeschen Drock vun enger CSV op Fonction publique verfall sin.** A dofir well CSV och léiwer héichdiploméiert employéén astellen, well do gët et keng objektiv Rekrutementskritären.

negativ neutral positiv

Saz iwwersprangen

Uleedung

Utilisateur: **username**
Schonns **0** Sätz bewäert.

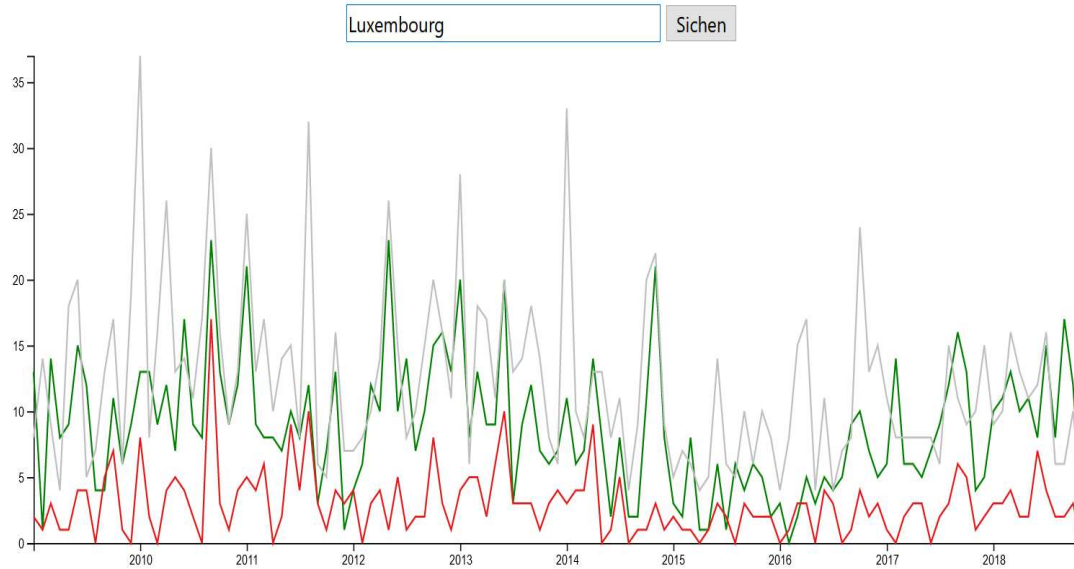
Dësen Tool ass Deel vum Projet STRIPS - A Semantic Search Toolbox for the Retrieve of Similar Patterns in Luxembourgish Documents.
Daniela Gierschek, Joshgun Sirajzade, Christoph Purschke, Peter Gilles, Christoph Schommer

STRIPS

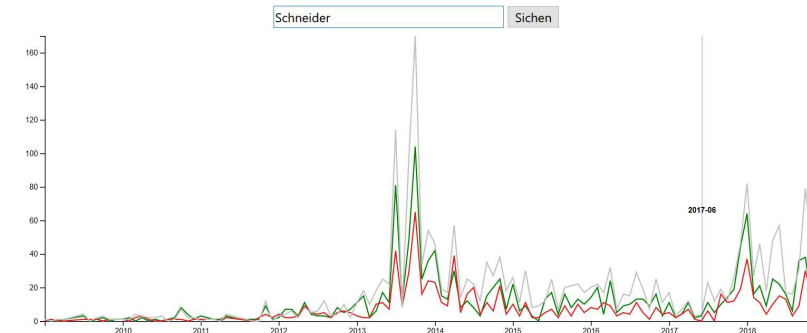
<http://sentilux.uni.lux/stripsannotation/sentences.php>

Joshgun Sirajzade
Christoph Schommer
Christoph Purschke
Peter Gilles
Daniela Gierschek

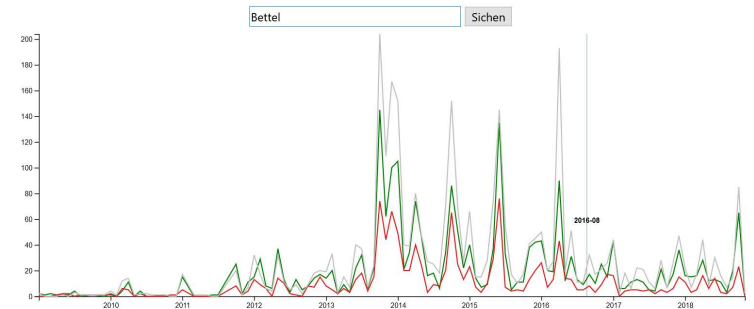
STRIPS Timeline Search



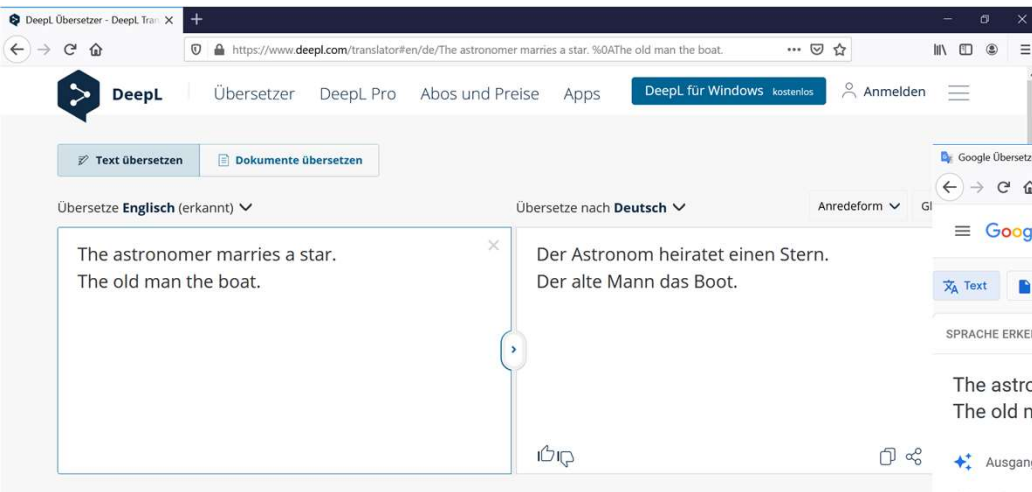
STRIPS Timeline Search



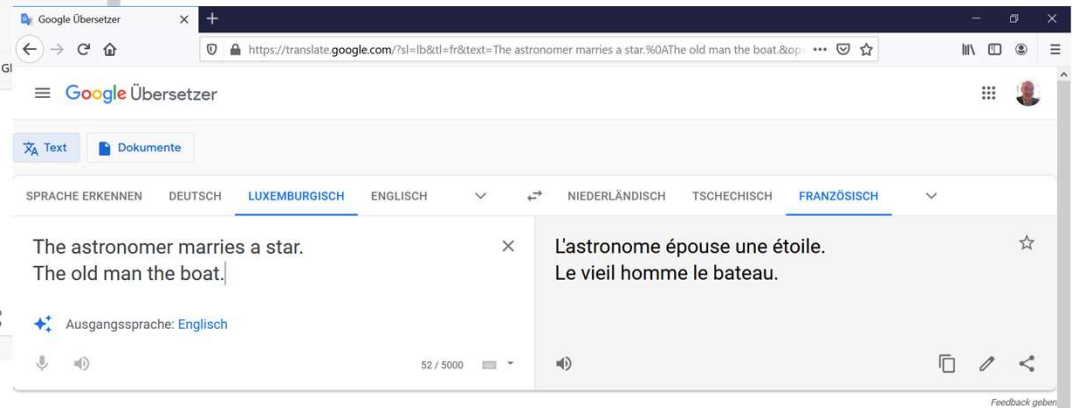
STRIPS Timeline Search



When NLP does not work



DeepL Pro — schnelle, präzise und sichere Übersetzungen



Verlauf



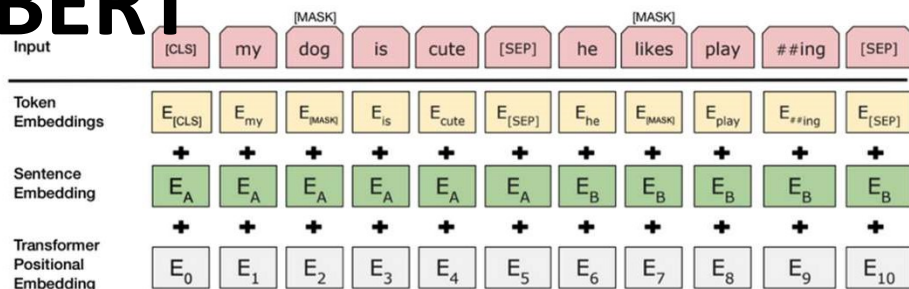
Gespeichert



Etwas beitragen

Next Generations... 3 examples

BERT



Source: [BERT](#) [Devlin et al., 2018], with modifications

wav2vec: Unsupervised Pre-training for Speech Recognition
Steffen Schneider, Alexei Baevski, Ronan Collobert, Michael Auli

We explore unsupervised pre-training for speech recognition by learning representations of raw audio. wav2vec is trained on large amounts of unlabeled audio data and the resulting representations are then used to improve acoustic model training. We pre-train a simple multi-layer convolutional neural network optimized via a noise contrastive binary classification task. Our experiments on WSJ reduce WER of a strong character-based log-mel filterbank baseline by up to 36% when only a few hours of transcribed data is available. Our approach achieves 2.43% WER on the nov92 test set. This outperforms Deep Speech 2, the best reported character-based system in the literature while using two orders of magnitude less labeled training data.

Subjects: [Computation and Language \(cs.CL\)](#)
Cite as: [arXiv:1904.05862 \[cs.CL\]](#)
(or [arXiv:1904.05862v1 \[cs.CL\]](#) for this version)

Submission history
From: Steffen Schneider [view email]
[v1] Thu, 11 Apr 2019 17:29:30 UTC (82 KB)
[v2] Fri, 24 May 2019 06:09:03 UTC (82 KB)
[v3] Tue, 2 Jul 2019 06:16:44 UTC (82 KB)
[v4] Wed, 11 Sep 2019 08:19:49 UTC (86 KB)

Download:
• PDF
• Other formats (license)

Current browse context: [cs.CL](#)
< prev | next >
new | recent | 1904
Change to browse by: [cs](#)

References & Citations
• NASA ADS
• Google Scholar
• Semantic Scholar

DBLP - CS Bibliography
[listing](#) | [bibtext](#)
Steffen Schneider
Alexei Baevski
Ronan Collobert
Michael Auli
[Export Bibtext Citation](#)

wav2net



A robot wrote this entire article. Are you scared yet, human?

GPT-3

GPT-3

We asked GPT-3, OpenAI's powerful new language generator, to write an essay for us from scratch. The assignment? To convince us robots come in peace

- For more about GPT-3 and how this essay was written and edited, please read our editor's note below





This is the last slide!

Thank you for
listening!