A Top Friendship: Measurement of $t\bar{t}H$ production in the H(bb) decay channel at ATLAS with Transformer Networks

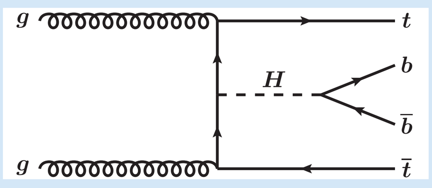
Yusong Tian (yusong.tian@cern.ch), on behalf of the ATLAS Collaboration
II. Physikalisches Institut, Georg-August-Universität Göttingen

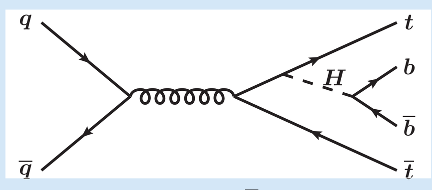
The $t\bar{t}H(H \to b\bar{b})$ Analysis in a Nutshell

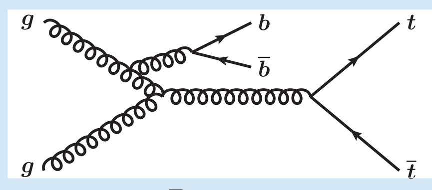
- Top quark heaviest $\Rightarrow$ strong Yukawa coupling
- $t\bar{t}H$: direct measurement of top Yukawa
- Subsequent $H \to b\bar{b}$ has largest branching ratio (58 %)

$t\bar{t}H$ interesting for new physics

Measure $t\bar{t}H(b\bar{b})$!


t-channel $t\bar{t}H(b\bar{b})$ production


s-channel $t\bar{t}H(b\bar{b})$ production


$t\bar{t} + b\bar{b}$ background

- **$t\bar{t}H(b\bar{b})$ still to be observed** - analysis measures signal strength $\mu = \frac{\sigma_{\text{measured}}}{\sigma_{\text{SM prediction}}}$
- **Simplified template cross-section (STXS) framework** $\Rightarrow$ measure in 6 STXS Higgs p_{T} regions the signal strength μ_i, to probe p_{T} dependent deviations from SM
- **Challenge**: large $t\bar{t}$ +jets background due to jets from b- or c- quarks
- Analysis done on 140 fb^{-1} of 13 TeV p-p data in 1 or 2 lepton (electron/muon) final states

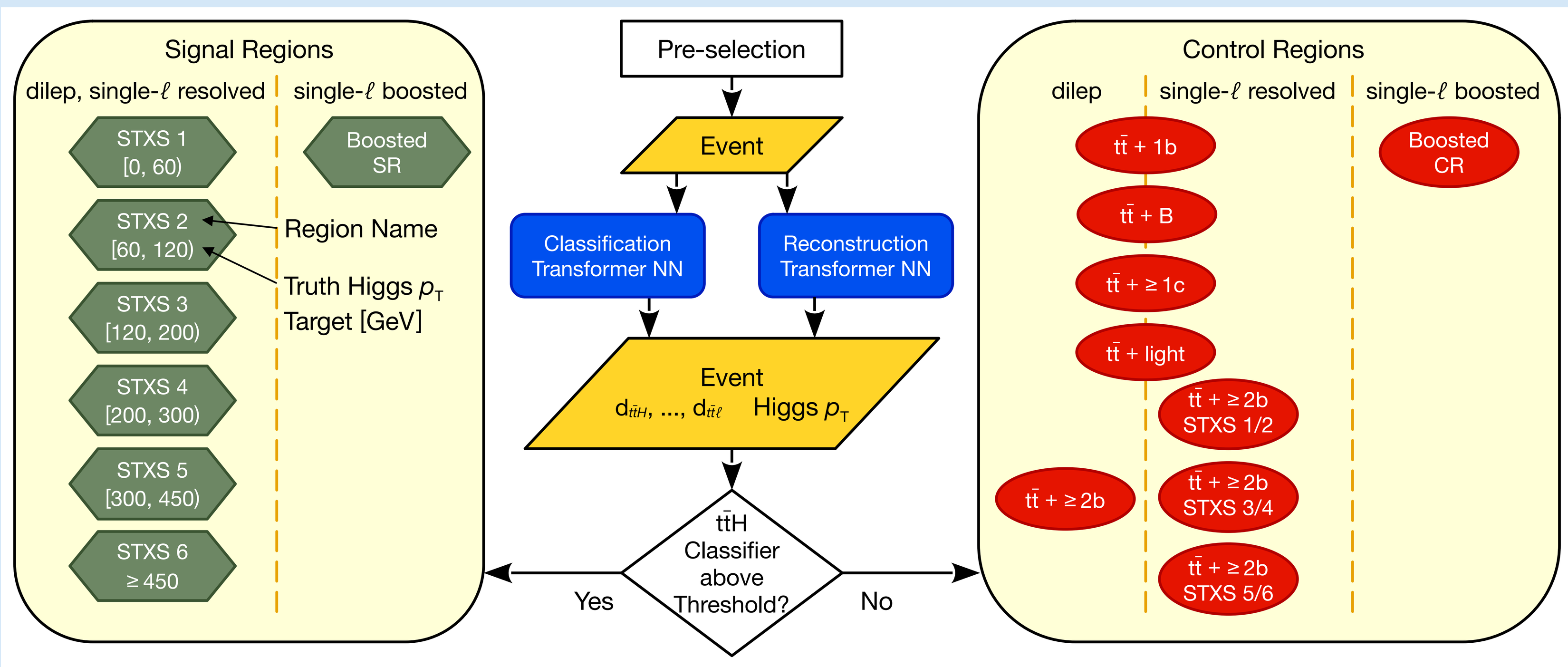
Improved multivariate analysis strategy over [1] using transformer NN for
- **Classification of signal and background events**
- **Higgs candidates reconstruction for p_T^H**

Event Classification and Higgs p_{T} Reconstruction

- Three channels: dilepton, single-lepton resolved, DNN-selected single-lepton boosted
- Pre-selection for events with certain number of jets and b-jets:

channel	# jets ($R = 0.4$) at WP,	in which # jets at WP	# jets ($R = 1.0$)
dilepton	3 at 85 %	2 at 70 %	
single-ℓ resolved	5	3 at 70 %	
single-ℓ boosted	4	3 at 85 %	1

Looser requirements than previous analysis: pre-selection efficiency for $t\bar{t}H$ 6.3 % - around 3 times higher



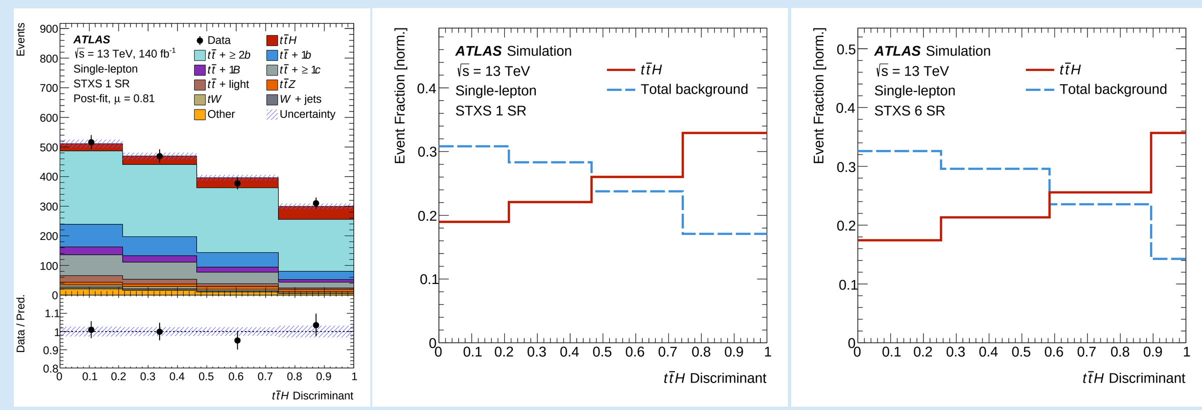
- Events separated into signal and control regions:
  - **Classification transformer** gives 6 discriminant values for each class: $t\bar{t}H$ signal, $t\bar{t} + 1b$, $t\bar{t} + 1B$, $t\bar{t} + \geq 2b$, $t\bar{t} + \geq 1c$, $t\bar{t}$ + light backgrounds

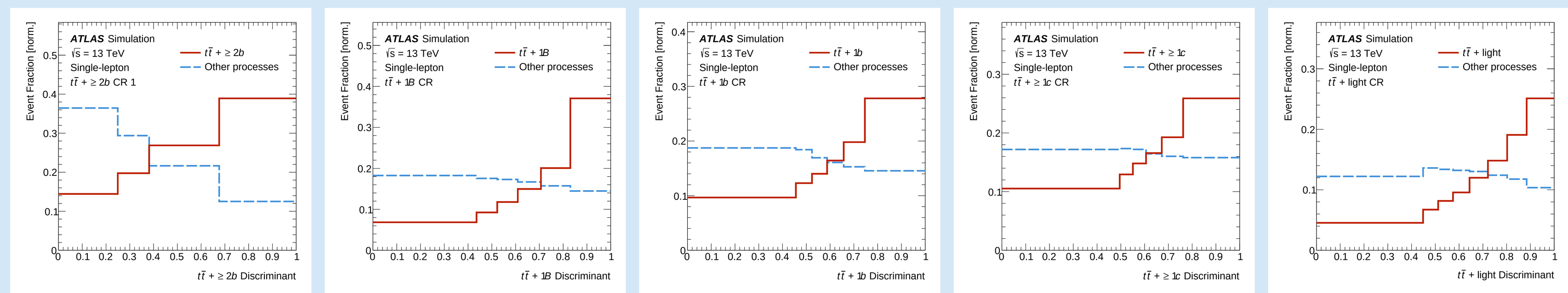
$$d_i = \frac{p_i}{\sum_{j \neq i} p_j \cdot \hat{N}_{ij}}, \text{ weights } \hat{N}_{ij} = \frac{N_j}{\sum_{k \neq i} N_k} \text{ from remaining classes} \quad (1)$$

  p: probability, N: expected event yield
  - Definition of discriminant function works well
  - Cut values on $d_{t\bar{t}H}$ for considering events signal-like: from $\max(\frac{S}{\sqrt{B}})$
  - Background grouping decided by the highest of the 5 d_i for bkg
  - **Reconstruction transformer** obtains p_T^H: predict which 2 jets most likely from H decay, get p_T^H from combining jet four-vectors
  Better performance than getting p_T^H from regression, whose predictions shift to central STXS bins

Signal and Background Classification Performance

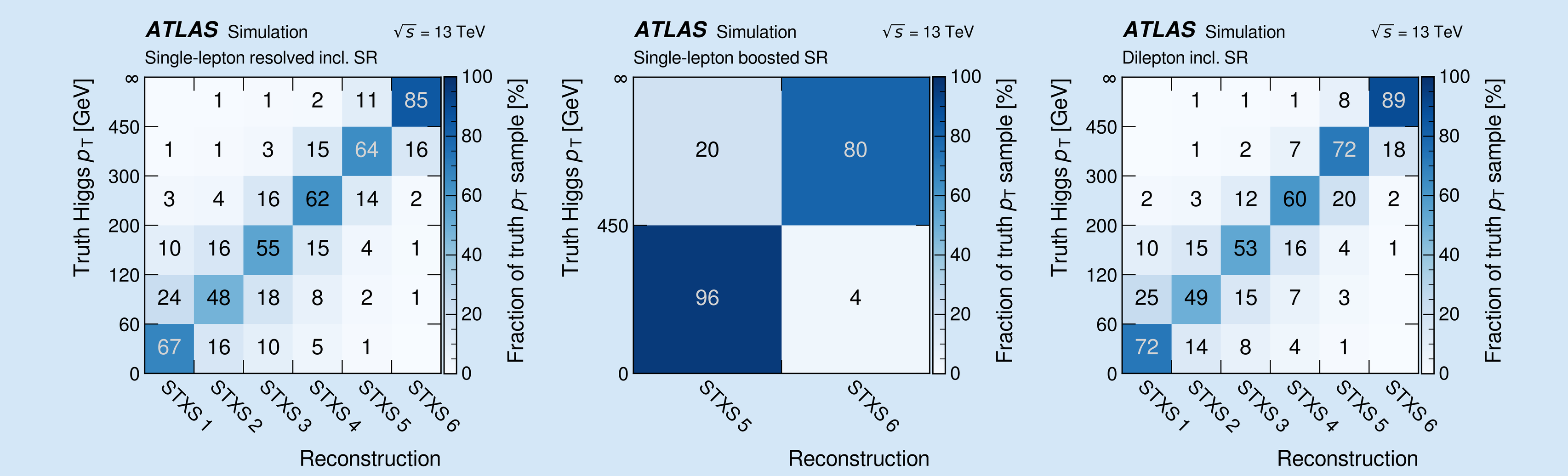
Data and MC comparison (single-ℓ), SR $\rightarrow$
Pre-fit (single-ℓ resolved) signal separation $\rightarrow$
and bkg classification $\downarrow$





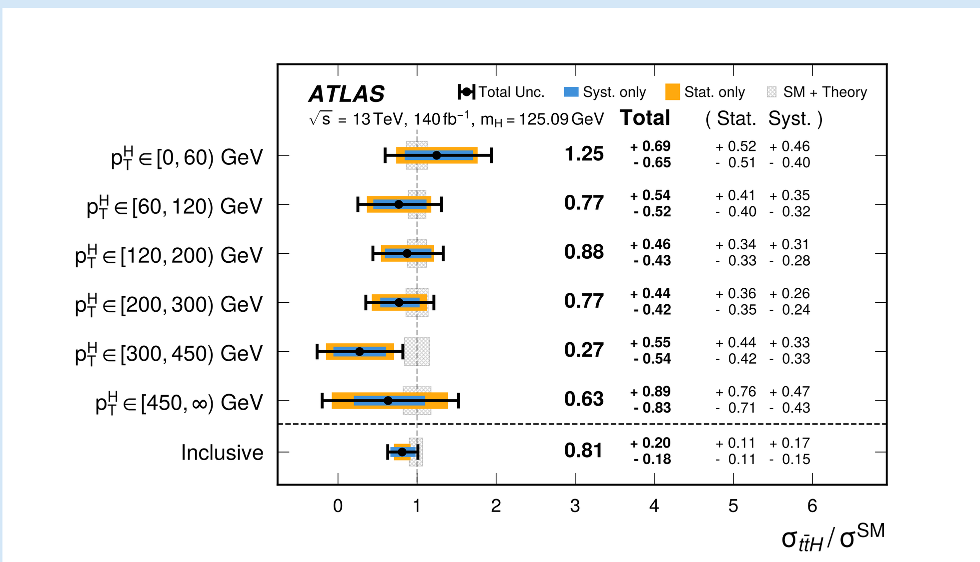
- The transformer NN approach in this analysis improved over the BDT approach as used previously [5]
- AUC 0.753 for single-lepton resolved and 0.774 for dilepton

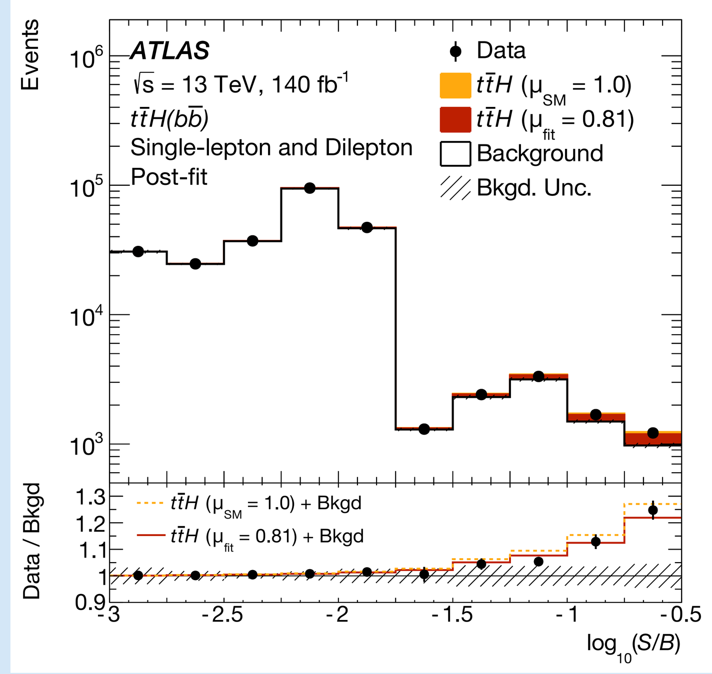
Higgs p_{T} Reconstruction Performance



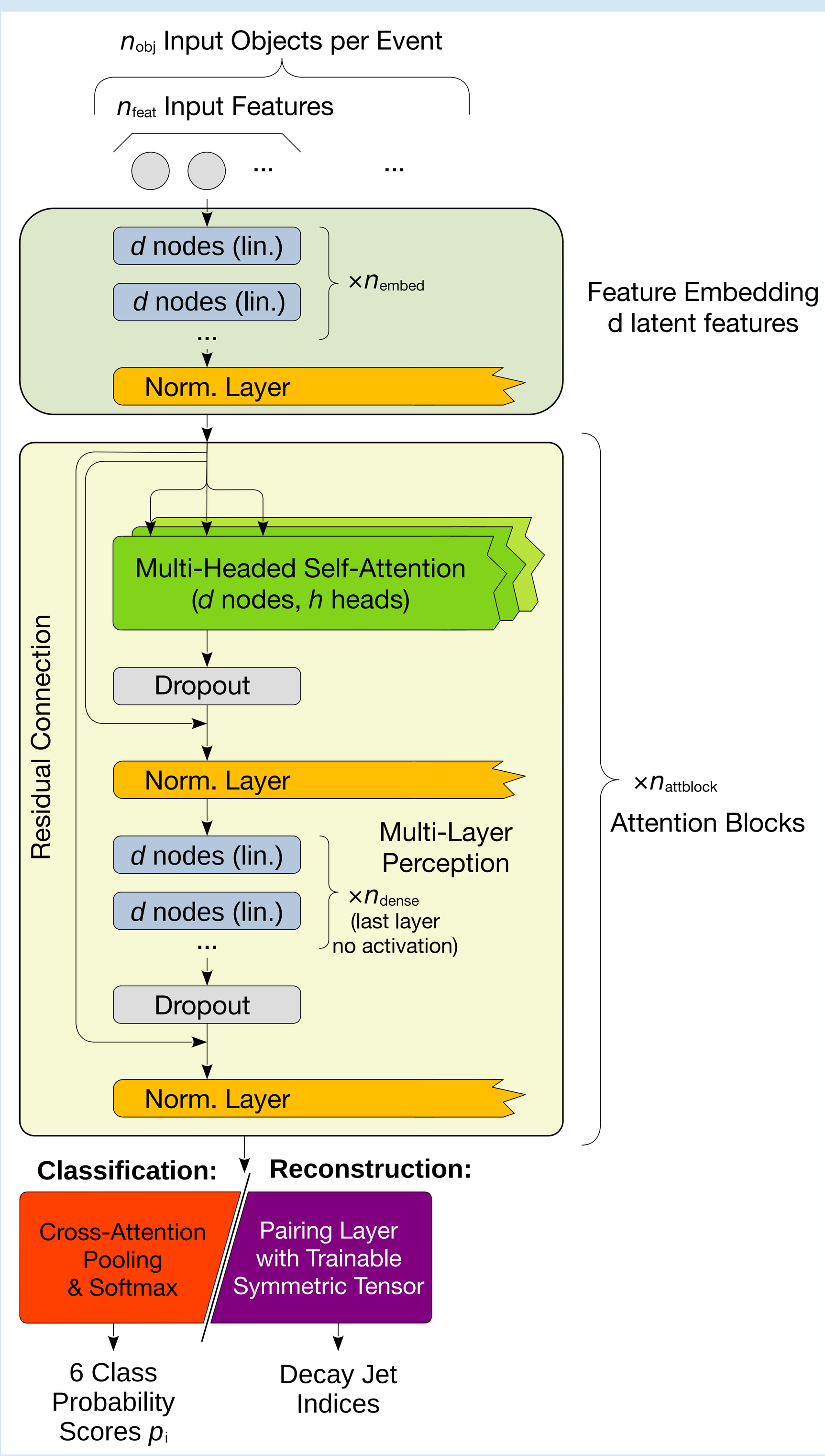
Results of the Analysis

- Excess of $t\bar{t}H(b\bar{b})$ events: observed (expected) significance 4.6 (5.4) std dev
- $t\bar{t}H$ signal strength for $m_H = 125.09$ GeV $\mu_{t\bar{t}H} = 0.81 \pm 0.11(\text{stat.})\,^{+0.20}_{-0.16}(\text{syst.})$, cross-section $\sigma_{t\bar{t}H} = 411 \pm 54(\text{stat.})\,^{+85}_{-75}(\text{syst.})$ fb, consistent with SM
- Measured in STXS bins





The Transformer Neural Networks



Transformer models like GPT [2] utilise attention mechanism [3]. Similar architecture used for $t\bar{t}H(b\bar{b})$ analysis
- Permutation-invariant architecture: ordering of input objects does not matter
- Uses: kinematic information of reconstructed electrons, muons, jets, E_T^{miss}; pseudo-continuous b-tagging score
- Trained on: classification - $t\bar{t}H$ signal, $t\bar{t}$ + jets bkg; reconstruction - only $t\bar{t}H$

Residual scheme modified from [3], where two separate residual connections are employed around the multi-head self-attention and the multi-layer perception sub-layers

Modified residual scheme results in:
- Less dependency on earlier blocks
- Slightly higher average classification accuracy

For p_T^H reco, pairing layer similar to [4]

Conclusions & Outlook

- Run 2 legacy analysis, significantly improved over the last analysis [1]
- Benefited from various improvements, including transformer NN used for MVA:
  - Define signal region and control regions
  - Higgs p_{T} reconstruction
- Discriminant function defined for event classification performs well
- Reconstructing jets from Higgs decay: better performance than obtaining p_T^H from regression

Paper: CERN-EP-2024-194
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HIGG-2020-24/

[1] ATLAS Collaboration, *Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP 06 (2022) 097

[2] T. Brown et al., *Language Models are Few-Shot Learners*, *Advances in Neural Information Processing Systems*, ed. by H. Larochelle et al., **33**, Curran Associates, Inc., 2020, 1877

[3] A. Vaswani et al., *Attention is all you need*, *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, Long Beach, California, USA: Curran Associates Inc., 2017, 6000

[4] M. J. Fenton et al., *Permutationless many-jet event reconstruction with symmetry preserving attention networks*, Phys. Rev. D 105 (2022) 112008

[5] ATLAS Collaboration, *Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D 97 (2018) 072016

42nd International Conference on High Energy Physics (ICHEP 2024), Prague, Czech Republic, 18.7.2024