

Τεχνητή Νοημοσύνη – Προκλήσεις και Ευκαιρίες

Ιωσήφ Σηφάκης | Αθήνα, 11 Μαρτίου 2026

- **Γέννηση και προσεγγίσεις:** Η τεχνητή νοημοσύνη ιδρύθηκε το 1956 με δύο ανταγωνιστικές προσεγγίσεις:
 - **Συμβολική** με βάση τη λογική και τις γλώσσες προγραμματισμού
 - **Εμπειρική** με βάση τη χρήση τεχνητών νευρωνικών δικτύων, που επικράτησε τελικά.
- **Εξέλιξη:**
 - **Κυκλική ανάπτυξη** με διακυμάνσεις μεταξύ υπερβολικών προσδοκιών και απογοητεύσεων, με **ελάχιστη επίδραση στην ψηφιακή** σύγκλιση μέχρι πρόσφατα.
 - **Επιτυχίες** από τις αρχές του αιώνα που οφείλονται στη **μηχανική μάθηση**, αρχικά σε **ανάλυση εικόνων** και αργότερα στην επεξεργασία φυσικών γλωσσών με τα **Μεγάλα Γλωσσικά Μοντέλα (LLM)**, π.χ. ChatGPT.

- **Όραμα και στόχοι:** Σήμερα, υπάρχει σύγχυση σχετικά με τον τελικό στόχο.
 - **Γενική Τεχνητή Νοημοσύνη (AGI):** Στόχος είναι να «ξεπεράσει τους ανθρώπους σε σχεδόν όλες τις γνωστικές λειτουργίες» χρησιμοποιώντας μόνο μηχανική μάθηση.
 - **Νοημοσύνη ανθρώπινου επιπέδου:** Στόχος είναι η κατασκευή μηχανών που κατανοούν τον κόσμο, συνδυάζουν γνώσεις και επιτυγχάνουν στόχους.

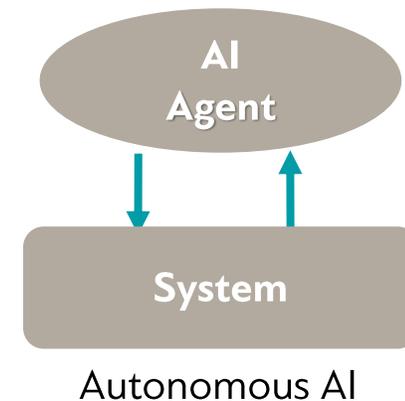
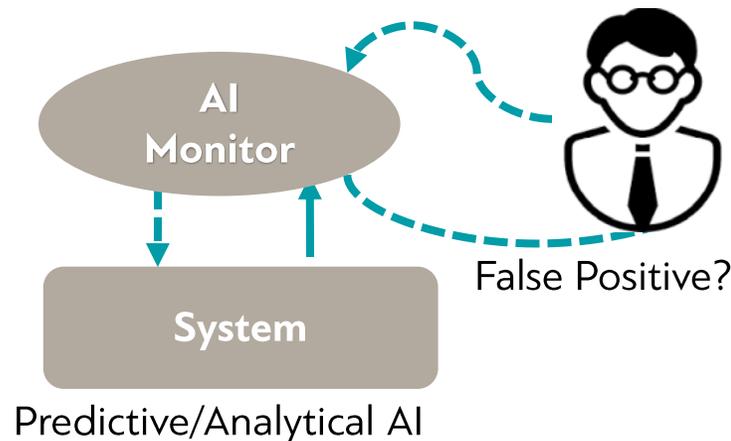
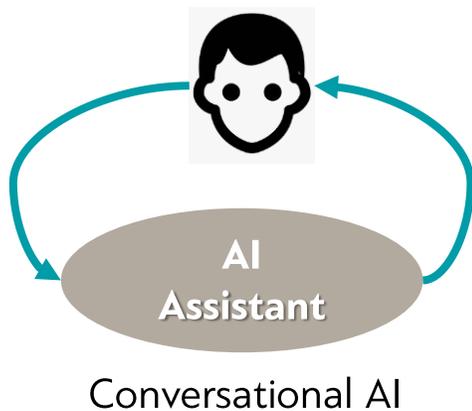
Διαφορά: Η ανθρώπινη νοημοσύνη δεν είναι απλά ένα σύνολο ετερόκλητων λειτουργιών -- συνδυάζει ορθολογικά οργανωμένη χρήση γνώσεων και στοχευμένη δράση.

□ Η ΤΝ βρίσκεται στα πρώτα της βήματα.

- Διαθέτει τα «δομικά στοιχεία», αλλά λείπουν οι αρχές για να τα συναρμολογήσουμε σε ολοκληρωμένα, αξιόπιστα συστήματα (όπως η κατασκευή μιας γέφυρας).
- Εστιάζει στους **βοηθούς** (assistant), που παρέχουν υπηρεσίες αλληλοεπιδρώντας με χρήστες, πχ ChatGPT.

□ Τρεις διαφορετικοί τρόποι χρήσης:

1. **Βοηθοί** που, σε αλληλεπίδραση με τον χρήστη, παρέχουν μια συγκεκριμένη υπηρεσία..
2. **Αναλυτές** συστημάτων που συνθέτουν γνώσεις για την ανίχνευση ή πρόβλεψη κρίσιμων καταστάσεων.
3. **Πράκτορες** (agents) ενός αυτόνομου συστήματος, π.χ. ο αυτόματος πιλότος ενός αυτοκινήτου.



Βασική πρόκληση: αυτόνομα συστήματα που παρέχουν **ασφαλείς υπηρεσίες χωρίς ανθρώπινη παρέμβαση.**

- ❑ **Τα αυτόνομα συστήματα:** αντικατάσταση της **ανθρώπινης εργασίας**, όπως προβλέπεται από το IoT, π.χ. αυτόνομα συστήματα οδήγησης, έξυπνα δίκτυα, έξυπνα εργοστάσια, έξυπνες πόλεις, αυτόνομα τηλεπικοινωνιακά συστήματα.
 - **Ατομική και συλλογική συμπεριφορά:** σύνολο από **πράκτορες** που επιδιώκουν **ατομικούς στόχους** ενώ παράλληλα συνεργάζονται επιδιώκοντας **κοινούς στόχους** για να επιτύχουν **συλλογική νοημοσύνη**.
 - **Προσαρμοστικότητα:** αντιλαμβάνονται, προβλέπουν, και προσαρμόζονται σε **δυναμικά περιβάλλοντα** και **διαχειρίζονται αντικρουόμενους στόχους**, όλα αυτά ενώ λειτουργούν συνεχώς και εξελίσσονται.

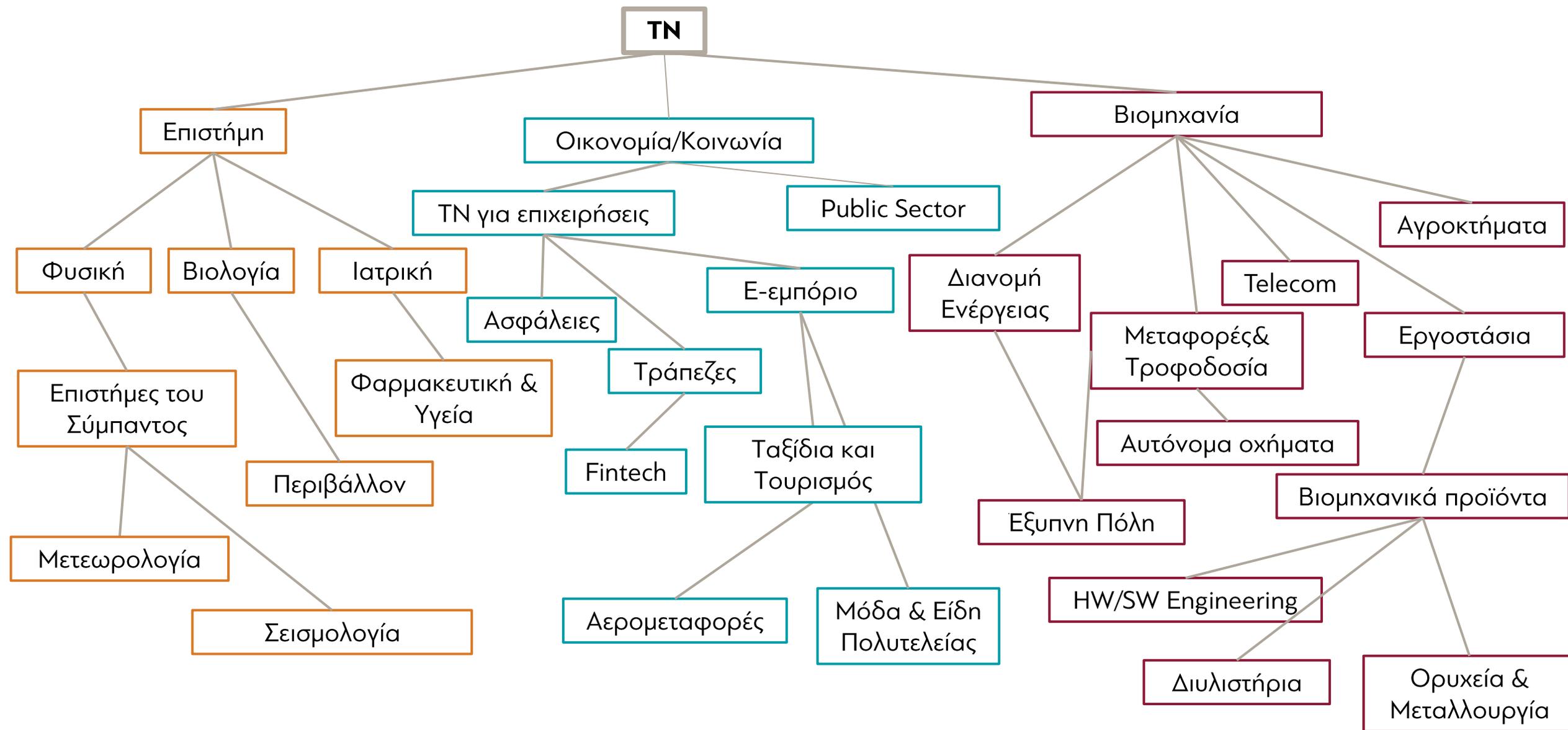
- ❑ **Το όραμα της αυτονομίας:** ενέχει μια σειρά **τεχνικών προκλήσεων** που υπερβαίνουν κατά πολύ τις δυνατότητες της τρέχουσας τεχνολογίας που στηρίζεται στη μηχανική μάθηση.
 - **Άστοχες προβλέψεις:** υπόσχονται πλήρη αυτονομία, άμεσα, π.χ. αυτόνομα αυτοκίνητα έως το 2020,
 - **Χαμηλή αξιοπιστία:** υπηρεσίες σε στατικά, ψηφιακά περιβάλλοντα, π.χ. φίλτρα ηλεκτρονικού ταχυδρομείου.

- ❑ **Οι προϋποθέσεις για την επιτυχία:**
 - **συλλογιστικές ικανότητες:** που να επιτρέπουν την ορθολογική επίτευξη στόχων (τεχνικών, νομικών, ηθικών)
 - **αξιοπιστία:** υψηλή για ασφαλή κρίσιμη λειτουργία (π.χ. $<10^{-8}$ αστοχίες/ώρα), πράγμα ανέφικτο σήμερα.

- **Η βασική διαφορά:** Τα συστήματα ΤΝ, από τη φύση τους, διαφέρουν από τα παραδοσιακά συστήματα:
 - **Αποτελεσματικά:** στη δημιουργία εμπειρικής γνώσης από αισθητηριακά ή γλωσσικά δεδομένα.
 - **Μη εξηγήσιμα:** η συμπεριφορά τους δεν περιγράφεται με μαθηματικά μοντέλα και είναι πρακτικά αδύνατο να εγγυηθεί κανείς την αξιοπιστία τους.

- **Το πρόβλημα της ασφάλειας:**
 - Η ασφάλεια των παραδοσιακών συστημάτων (ανελκυστήρες, αεροπλάνα) βασίζεται σε **μαθηματικές μεθόδους**
 - Η **μη εξηγησιμότητα** της ΤΝ καθιστά αυτές τις δοκιμασμένες μεθόδους μη εφαρμόσιμες.

- **Το πρόβλημα των γνωστικών ιδιοτήτων:** Για να είναι αξιόπιστη η ΤΝ, πρέπει να ικανοποιεί **ανθρώπινες γνωστικές ιδιότητες** (υπεύθυνα, ευθυγραμμισμένα, ηθική τεχνητή νοημοσύνη).
 - **Ελλειψη επιστημονικής βάσης:** Οι τρέχουσες μελέτες σχετικά με αυτές τις ιδιότητες στερούνται επιστημονικής ακρίβειας.
 - **Δυσκολίες:**
 - **Απροσδιοριστία της γνώσης:** Δεν μπορούμε τεχνικά να διακρίνουμε μεταξύ ενός συστήματος που ψεύδεται ή που απλώς δεν γνωρίζει (π.χ. «Η γη είναι επίπεδη»).
 - **Αξιακά κριτήρια:** Πώς μετράμε την ικανότητα ενός συστήματος να εσωτερικεύει και να ενεργεί με βάση τις **ανθρώπινες αξίες** (π.χ. Τι διακρίνει πραγματικά έναν οδηγό ΤΝ από έναν ανθρώπινο οδηγό;).



- ❑ **Ο ανθρώπινος νους:** Διαθέτει **κοινή λογική**, ένα σημασιολογικό μοντέλο του κόσμου —δυναμικό δίκτυο γνώσεων που χτίζεται και ενημερώνεται καθ' όλη τη διάρκεια της ζωής μέσω της μάθησης και της λογικής.
- ❑ **Ανθρώπινη κατανόηση:** Συνδυάζει δύο διαδρομές:
 - **Από κάτω προς τα πάνω:** Αισθητήρες που τροφοδοτούν το νοητικό μοντέλο.
 - **Από πάνω προς τα κάτω:** Το νοητικό μοντέλο που ελέγχει αισθητήρες και ενεργοποιητές.

❑ **Ενα παράδειγμα:**

- **Άνθρωπος:** Αναγνωρίζει μια πινακίδα STOP καλυμμένη με χιόνι, επειδή η εικόνα ταιριάζει με το υπάρχον **εννοιολογικό μοντέλο** (μέγεθος, χρώμα, σχήμα, θέση).
- **Μηχανή:** Ενα σύστημα ΤΝ πρέπει να εκπαιδευτεί σε **κάθε πιθανή παραλλαγή** (χιόνι, ομίχλη, γκράφιτι) για να επιτύχει το ίδιο αποτέλεσμα.



❑ Ανάπτυξη της επιστημονικής γνώσης:

- **Παραγωγή πληροφορίας/γνώσης** από πολυδιάστατα δεδομένα, ξεπερνώντας τις ανθρώπινες δυνατότητες σε τομείς όπως η βιολογία, η ιατρική και η επιστήμη των υλικών.
- **Βοήθεια στην αντιμετώπιση** των πλανητικών προκλήσεων (κλίμα, πόροι, καταστροφές) επιτρέποντας την **προβλεψιμότητα**, αλλά προσφέροντας μόνο **περιορισμένη κατανόηση**.

Κανόνας: Οι άνθρωποι πρέπει να διατηρήσουν τον **έλεγχο της παραγωγής επιστημονικής γνώσης**, με τις μηχανές να λειτουργούν ως συνεργάτες.

❑ Αυτοματοποίηση, παραγωγικότητα και ποιότητα ζωής:

- **Βελτίωση της ποιότητας ζωής:** καθώς αυξάνονται η αξιοπιστία και οι δυνατότητες της ΤΝ.
- **Παραγωγικότητα:** απλοποίηση και αυτοματοποίησή σύνθετων διαδικασιών, ταχεία και βέλτιστη λήψη αποφάσεων, φιλικές προς τον χρήστη υπηρεσίες (υγειονομική περίθαλψη, εκπαίδευση, μεταφορές).

❑ Κανόνες:

- **Ο βαθμός αξιοπιστίας** μιας μηχανής να είναι ανάλογος με την κρισιμότητα των αποφάσεών της.
- **Ασφαλής ανθρώπινη παρέμβαση** όταν ένα αυτόνομο σύστημα σφάλλει.

Μια Ματιά στο Μέλλον – Συστημικές Επιπτώσεις της ΤΝ

□ Γνωστική & Κοινωνική Δομή:

- **Νοητική Ατροφία:** Η ανάθεση εργασιών αποδυναμώνει την ανθρώπινη συγκέντρωση, κατανόηση και δημιουργικότητα:
- **Σχέσεις:** Αναδιαμόρφωση των οικογενειακών και κοινωνικών σχέσεων.

□ Οικονομία και διακυβέρνηση:

- **Αγορά εργασίας:** Απαιτεί μαζική αναβάθμιση των δεξιοτήτων του εργατικού δυναμικού – ανεργία
- **Κρατική εξουσία:** Εξασθενεί στη λήψη αποφάσεων:
 - **Ιδιωτικοποίηση** των υπηρεσιών και υποδομών για την διαχείριση της γνώσης και της πληροφορίας,
 - **Επιρροή** των μεγάλων τεχνολογικών εταιρειών...

□ Αναδιαμόρφωση της διεθνούς τάξης και της γεωπολιτικής σκηνής:

- **Διάχυση εξουσίας:** Αποδυνάμωση των εθνικών κρατών, Ιδιωτικός έλεγχος κρίσιμων υποδομών.
- **Ανατροπή της μεταπολεμικής τάξης:** Αλλάζουν οι έννοιες της σύγκρουσης, αποτροπής και υπεροχής,
- **Αγώνας εξοπλισμών:** Υπεροχή στην ταχύτητα λήψης αποφάσεων παρά στην καταστροφική δύναμη.
- **Ευθύνη αποφάσεων:** Μετατόπιση από τους ανθρώπους στις μηχανές.
- **Ασαφή όρια:** Μεταξύ ειρήνης/πολέμου, πολιτικών/στρατιωτικών στόχων και φυσικού/ψηφιακού κόσμου

Συμπεράσματα – Η Ευθύνη μας για το Μέλλον

□ Ενα ιστορικό ορόσημο:

- Σε αντίθεση με άλλες τεχνολογίες (πυρηνική ενέργεια, γενετική μηχανική), η ΤΝ είναι μοναδική.
 - ανοίγει **νέους ορίζοντες** για τη γνώση,
 - ενώ ταυτόχρονα **θέτει σε κίνδυνο** το ανθρώπινο πλεονέκτημα του να ορίζουμε την ύπαρξή μας,
 - έχει **βαθιά επίδραση** στα άτομα, την οικονομία και την κοινωνία.
- Τα **οφέλη είναι ανάλογα με τους κινδύνους**, γεγονός που τροφοδοτεί έντονες συζητήσεις.

□ Η ανάγκη για ολιστικό έλεγχο του ρίσκου και κανονιστικά πλαίσια:

- Να αποφύγουμε **στενές προσεγγίσεις** που εστιάζουν μόνο στο δίλημμα «αποτελεσματικότητα έναντι ασφάλειας».
- Να εξετάσουμε **την ΤΝ σε όλες τις μορφές**, τις εφαρμογές και τις μελλοντικές προοπτικές της.
- Να αναζητήσουμε **διεθνή κανονιστικά πλαίσια** για την χρήση της ΤΝ, δεδομένης της διαδεδομένης και ευρείας φύσης της ΤΝ, η οποία υπερβαίνει τα γεωγραφικά και εθνικά σύνορα

□ Τελική σκέψη:

- **Να μεγιστοποιήσουμε τα οφέλη** με γνώμονα το κοινωνικό όφελος και την προστασία θεμελιωδών αξιών.
- Το «**πρόβλημα**» **δεν είναι η ΤΝ** αλλά η αδυναμία μας να ελέγξουμε και να ρυθμίσουμε την χρήση της



Ευχαριστώ