Tutorial: Theory of Unsupervised Deep Learning, Harold Szu

Fellows of IEEE, AIMBE, OSA, INNS, SPIE; Research Ordinary Prof. BME, Catholic Univ. Am. Wash DC;
A Founder of INNS 1988, Foreign Academician of Russian Academy of NL Sciences#135, Jan 15, 1999 St. Petersburg;
Senior Scientist, Army Night Vision ESD, Ft. Belvoir, VA

- **Synopsis: NI** is defined to be **Unsupervised/Supervised Deep Learning (U/S DL) taking two (Neurons & Glial Cells) to Tango** for Darwinian Survival if on if. **2 conditions:** (1) **Homeostasis** a al **Herman Helmholtz Min Free Energy (MFE)/Least Mean Squares (LMS)** at Const. Temp. (general than **SDL LMS**) (2) **Power of Pair**. When sensors stimulus brainwaves and pair **coincidence** MFE: "agree is signal & kept, and disagree are noises thus rejected, relaxing towards MFE (Homospiens at 37oC for optimum elasticity of hemoglobium; chicken @40oC hatching egg)

- If we know the shortfall, we know how to do better. If we follow the principle of **Physics (Entropy as a measure of uniformity for MFE) & Physiology (Glial Biology for unsupervised)**, the breakthrough of 6 Glial cells from brain tumor **Glioma** to Alzheimer, Epileptic Seizure etc. disorders, to understand missing half of Einstein's, because Homosapiens brains have 3 lb weight with 10 B gray matter neurons & 100 B white matter Glial Cells.

 For the first we derived Glial Cells, $g_i \equiv -\Delta H/\Delta D_i$, where H is homeostasis MFE and D_i is dendrite net weighted sum of i-th neuron. Such a BNN take 10 B neurons &100 B Glial cells to dance Tangle unsupervised .

- **What do you expect to learn** from this **Seminar** (if 1-way) or **Colloquium** (if 2-way dialogue)

- **0. Courage (answer your own question), Comprehensive (Maxwell displacement current in capacitor), Complementarity (Q.M. commutator), Conscientious (Edison hard working); 10 group creativity rule how to be 1+1 =11, otherwise 1+1<1. ("Lesson in Creativity," Harold Szu, Ron Driggers, Appl. Opt. V.25, Aug. 2015)**
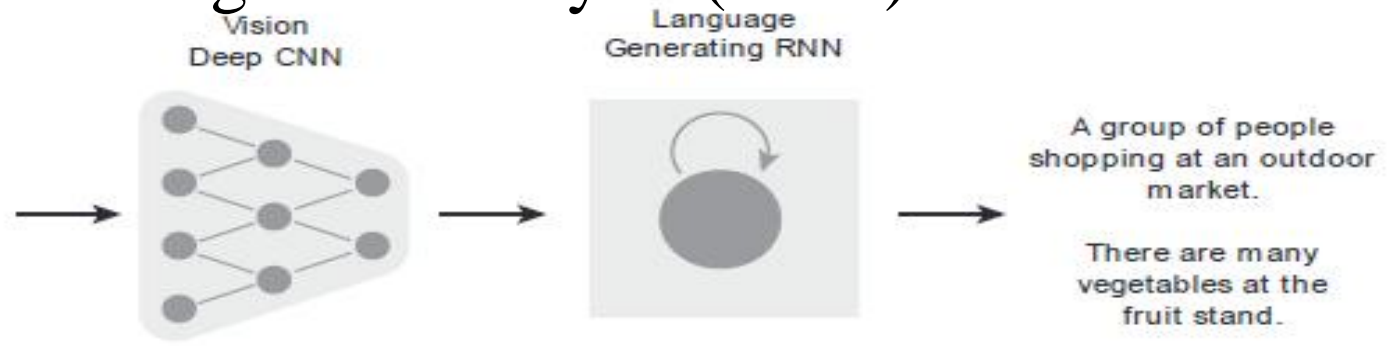
- **1. To Be Graduated**: Plenty jobs available at Internet Giants, Google Brain Alpha-Go (winning chess games), YouTube (discovery of the most favorite video is cat), as well as Facebook (independent of age, expression). **Alan Turing definition of Artificial Intelligence(AI) that can not tell whether it is human or computer**.

- **2. For Ph. D. Thesis**: We can compare AI Supervised Deep Learning (SDL) using ANN versus NI Unsupervised Deep Learning (UDL) BNN(**100 ways to skin a Cat**) ; **3. For faculty**: Seeking shortfalls to matching federal funding (beyond CR): **(DoC)** Consumer Index Machine IQ defines from Zoombot to House Maid; **(DoT)** Automated pilot i-Cars; **(DoD)** UXS swarming; **(DoE)** Robotic Rescue Missions in space, undersea, reactors.

Internet Google, Face Book, YouTube revive AI ANN for Big Data Analysis(BDA)

- *What is AI? Alan Turing introduced AI so that "one cannot tell the difference whether the other end of computer terminal is human or machine," circa WWII.*

- *When & How suddenly is the renaissance of Artificial Neural Net (ANN) Deep learning (DL)? Now the other end of computer can beat human in a chess game, face recognition , video perception.*

- **Google** co-founder *Sergey Brin* sponsored *AI AlphaGo* was surprised by the intuition, the beauty & the communication skills displayed by AlphaGo (阿尔法围棋) **which** had beat Korean Go Grand Master *Lee Sedol (李世石) in* 4 to 1 score on March 9-15, 2016, as millions watched on Internet; (第一局3月9日、第二局3月10日、第三局3月12日、第四局3月13日、第五局3月15日，每局在北京时间12点开赛)**.**

- **YouTube** has applied **Deep Learning** to annotate videos automatically & discovered the favorite video on YouTube turns out to be about favorite pet Cats.

- **Facebook** applied **Deep Learning wishes to achieve** aging & e-IQ independent facial expressions

Big Data Analysis (BDA)

Vision
Deep CNN

Language
Generating RNN

A group of people shopping at an outdoor market.

There are many vegetables at the fruit stand.

Yann LeCun (Facebook AI, NYU), Yoshua Bengio (Montreal) , Geoffrey Hinton (Google, Toronto) . "Deep Learning", Nature 521, pp. 436-444,May 28, 2015

A woman is throwing a frisbee in a park.

A dog is standing on a hardwood floor.

A stop sign is on a road with a mountain in the background

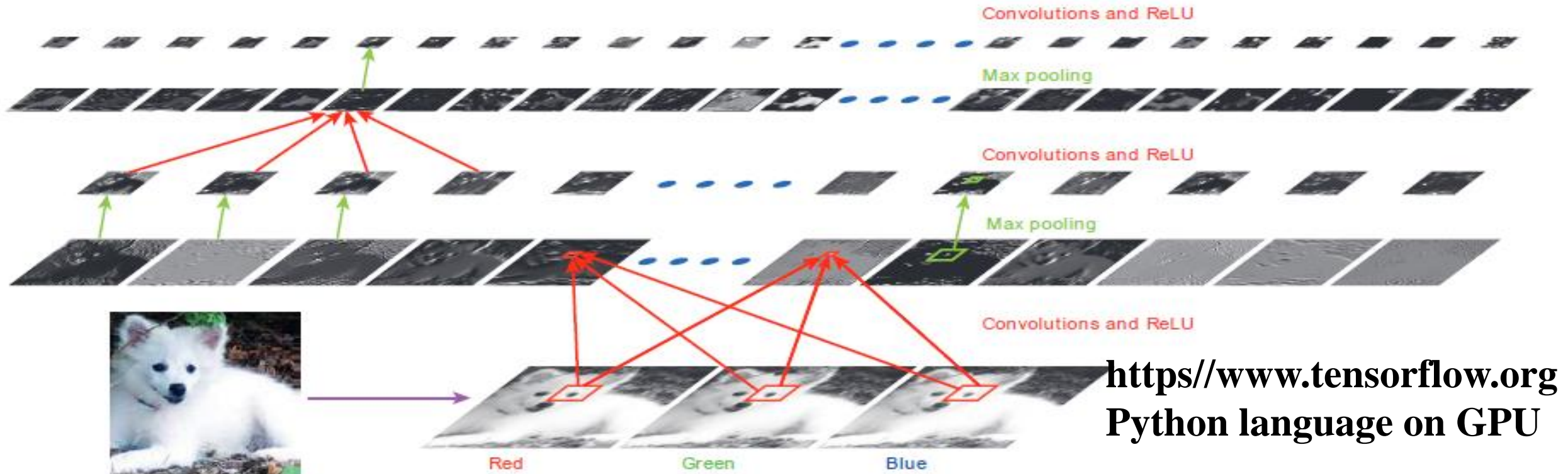A little girl sitting on a bed with a teddy bear.

A group of people sitting on a boat in the water.

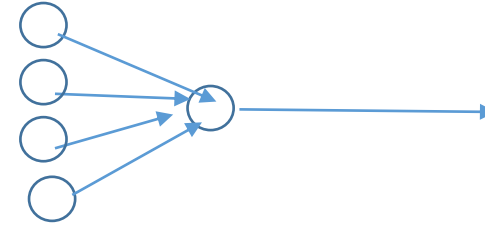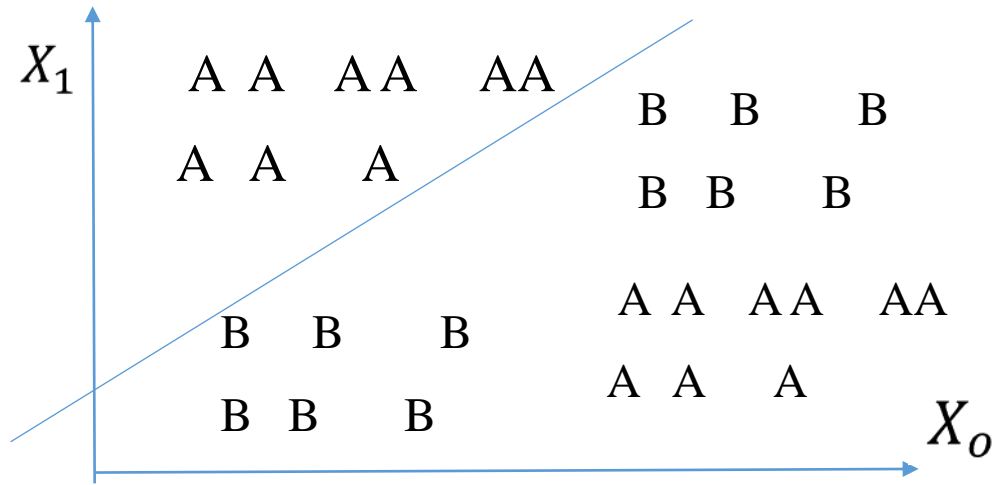A giraffe standing in a forest with trees in the background.

Training Procedure:

- **Input** has 4 layers edge at layer V1, curvature layer V2, texture layer v3, RGB color pixels per neurons at V4 , hidden layers Internal knowledge representation.
- **Hidden layers** has 10 to 100 layers.
- **Output** is a category vector has components, one for each class of object: cat, fox, dog, etc. whose numerical value will be given by a training supervisor.

Convolutions and ReLU

Max pooling

Convolutions and ReLU

Max pooling

Convolutions and ReLU

https//www.tensorflow.org
Python language on GPU

Red   Green   Blue

Why Deep (Multiple Layers) Learning (DL)?
We begin with a single layer ANN:

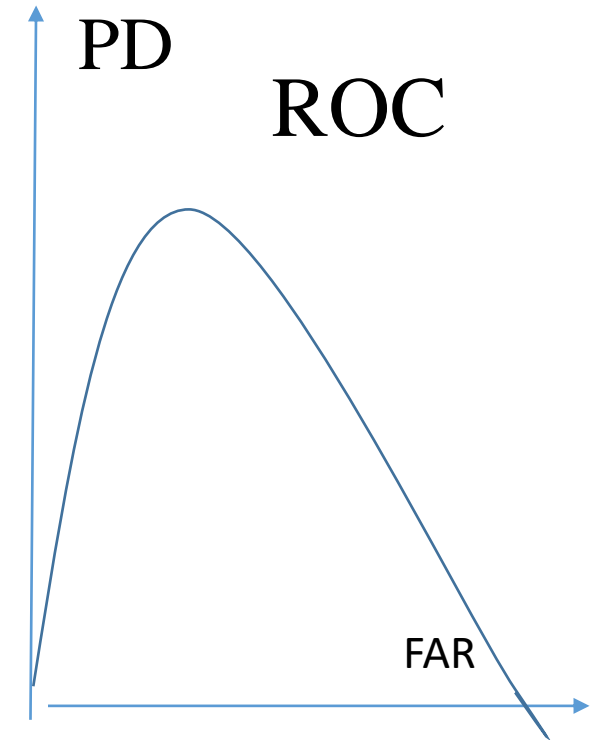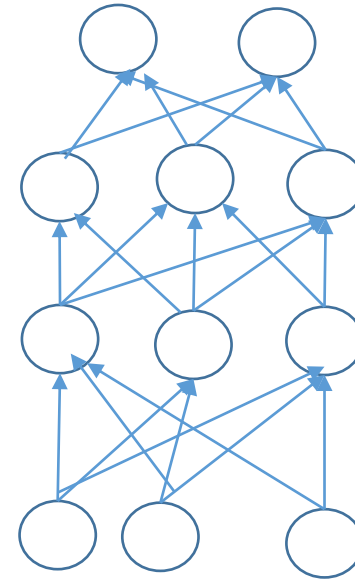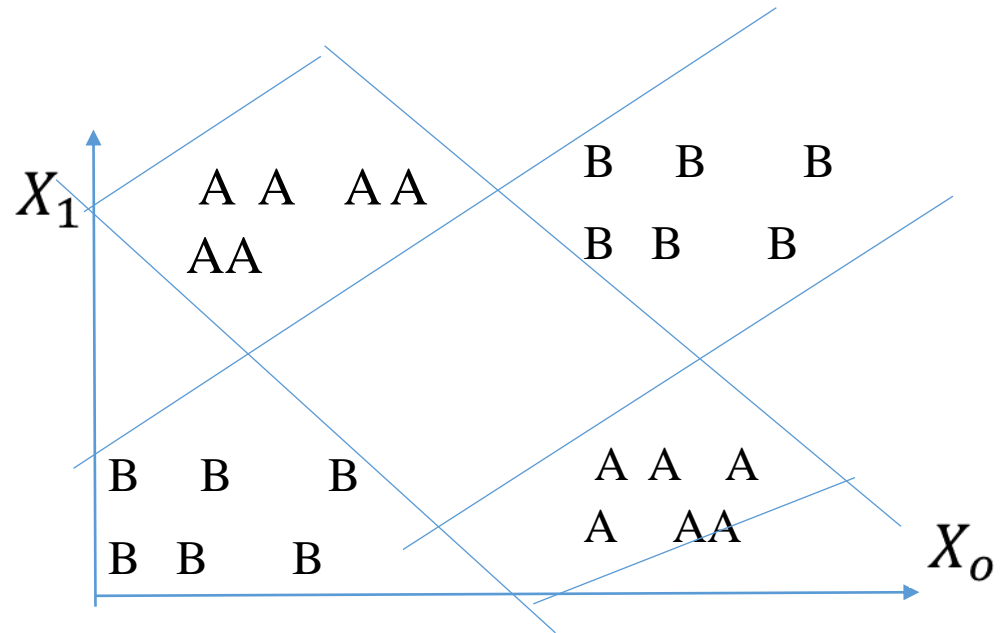$$S_j = \sigma(\sum_i [W_{j,i}]X_i - \theta)-$$

$$X_1 = \frac{-W_o}{W_1} X_o + \frac{\theta}{W_1}$$

1-layer Artificial Neural Network (ANN)
 is a **Linear Classifier** for target **A, e.g.** malignant cancer & non-target **B, e.g.**
benign cancer, with high **False Alarm Rate (FAR)**

Why do we need Multiple Layers (ML) Feature Extraction (FE) and Multiple Layers (ML) Machine Classifier?
Such ML ANN enjoys Convex Hulls Classifier when N> 3

- N-layers *ANN DL* are for a Convex Hull Classifier
- Enhance **Receiver Operation Characteristics (ROC),**
- Learning to Increase the **Probability of Detection (PD)** reduce the **False Alarm Rate (FAR)**,
- This goal *has not yet been directly mapped* into **Unsupervised Deep Learning**. (cf. **NI & BNN**; 2 Ph. D. Thesis: One in Data Mining, the other in Intelligent Robots)

a

Input
(2)

Hidden
(2 sigmoid)

Output
(1 sigmoid)

b

z

$$\Delta z = \frac{\partial z}{\partial y}\,\Delta y$$

$\frac{\partial z}{\partial y}$

$$\Delta y = \frac{\partial y}{\partial x}\,\Delta x$$

y

$\frac{\partial y}{\partial x}$

$$\Delta z = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}\Delta x$$

x

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}$$

c

Output units

$$y_l = f(z_l)$$

$$z_l = \sum_{k\,\varepsilon\,H2} w_{kl}\,y_k$$

Hidden units H2

$$y_k = f(z_k)$$

$$z_k = \sum_{j\,\varepsilon\,H1} w_{jk}\,y_j$$

Hidden units H1

$$y_j = f(z_j)$$

$$z_j = \sum_{i\,\varepsilon\,\text{Input}} w_{ij}\,x_i$$

Input units

d

Compare outputs with correct
answer to get error derivatives

$$\frac{\partial E}{\partial y_l} = y_l - t_l$$

$$\frac{\partial E}{\partial z_l} = \frac{\partial E}{\partial y_l}\frac{\partial y_l}{\partial z_l}$$

$$\frac{\partial E}{\partial y_k} = \sum_{l\,\varepsilon\,\text{out}} w_{kl}\frac{\partial E}{\partial z_l}$$

$$\frac{\partial E}{\partial z_k} = \frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial z_k}$$

$$\frac{\partial E}{\partial y_j} = \sum_{k\,\varepsilon\,H2} w_{jk}\frac{\partial E}{\partial z_k}$$

$$\frac{\partial E}{\partial z_j} = \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial z_j}$$
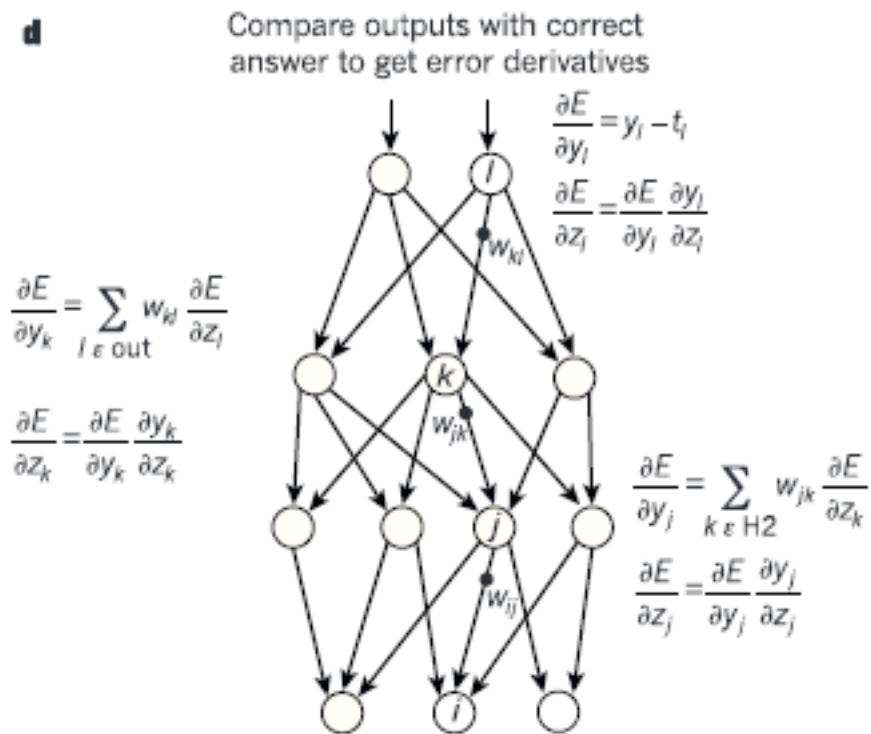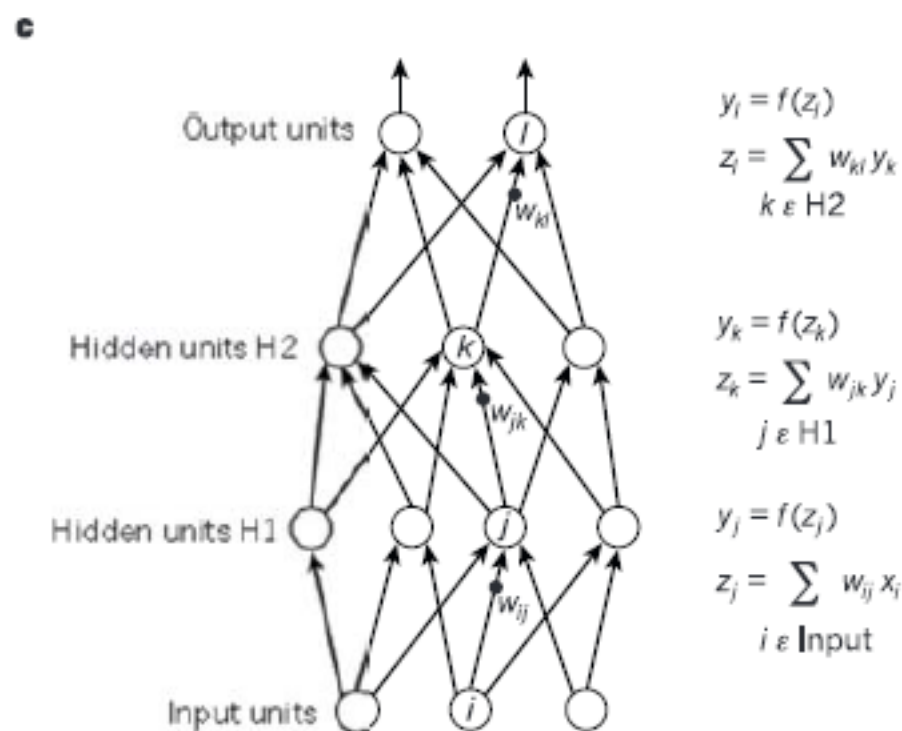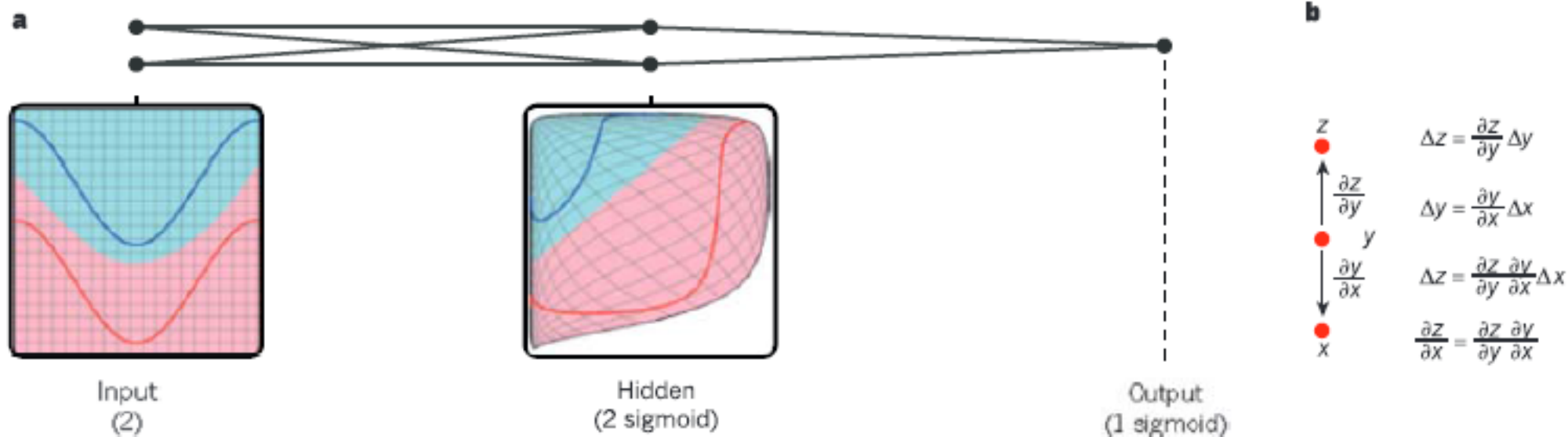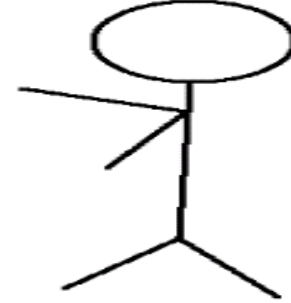
Introduction: Man-Machine Interface becoming important for peaceful co-existance
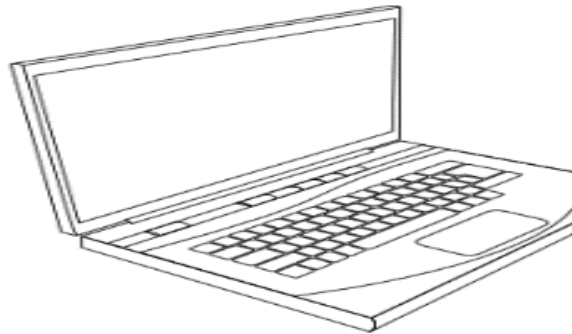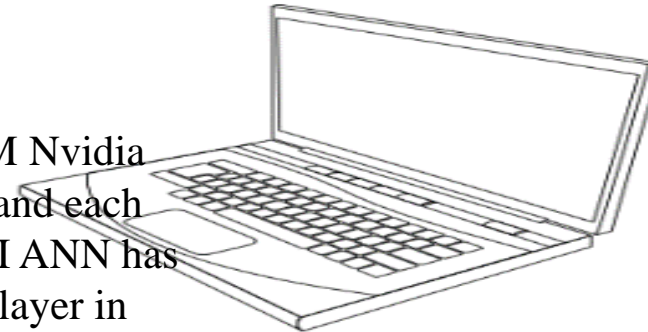

Why can't the machines learn new things quickly?
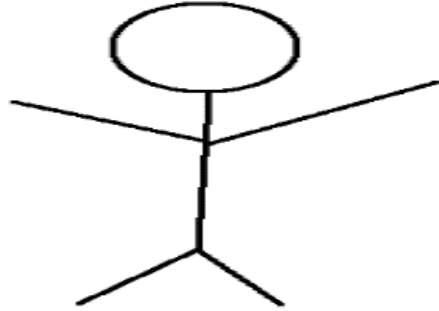
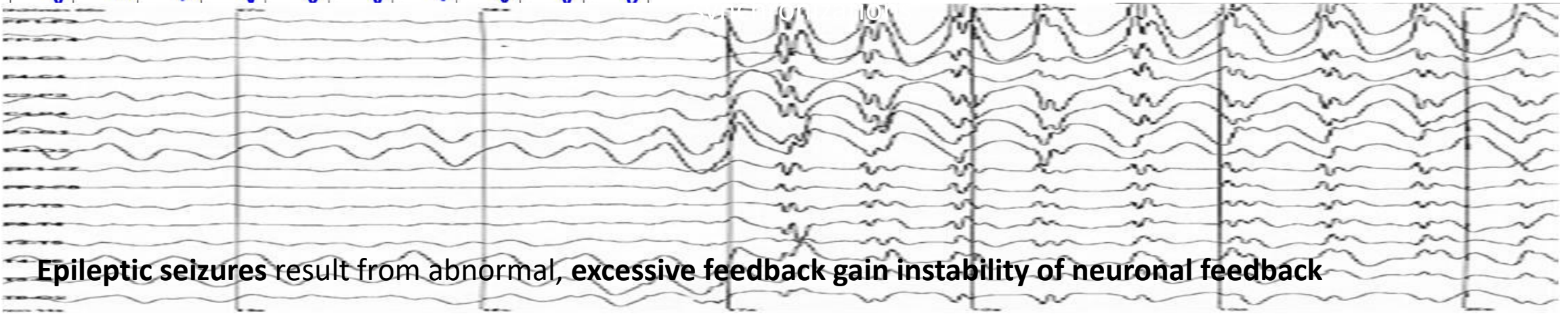Well, because they are just machines!

Why can't humans be as efficient as us?

Well, because they are just humans!

High Performance Computers, e.g. $M Nvidia made of 64 racks and each rack has 8 GPU. AI ANN has 100 layers & each layer in pseudo-real time

Biological Neural Networks—understanding brain waves due to neuron spiking population & local field potential for an early warning of the on-set [9] of Epileptic Seizure as 2nd order phase transition wearing f-EEG hat

neuronal spiking

local field potential

Right hand side image reveal the biological neural networks having axon linkages facilitating strong feedback

Epileptic seizures result from abnormal, **excessive feedback gain instability of neuronal feedback**
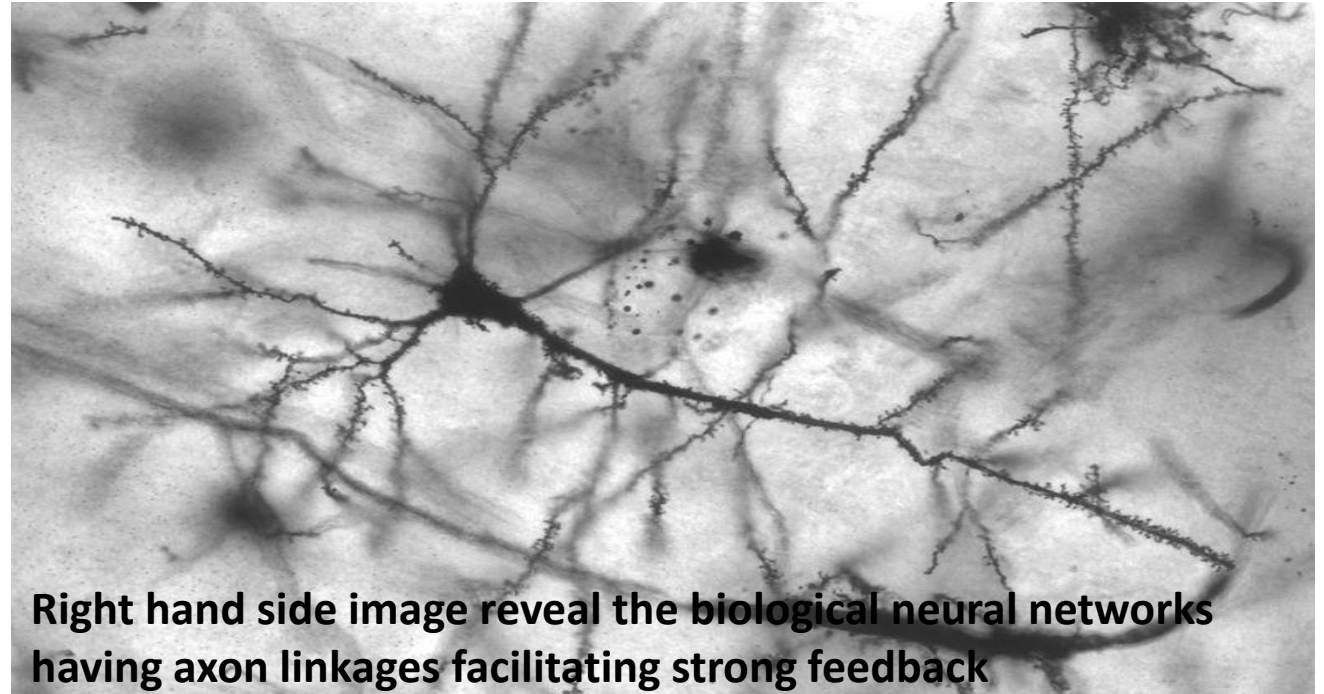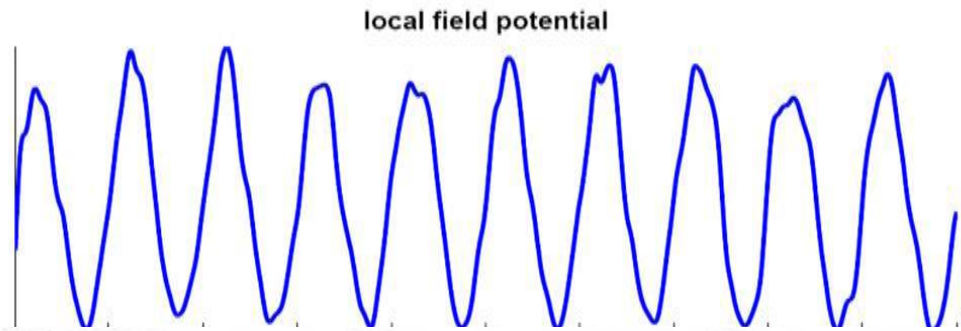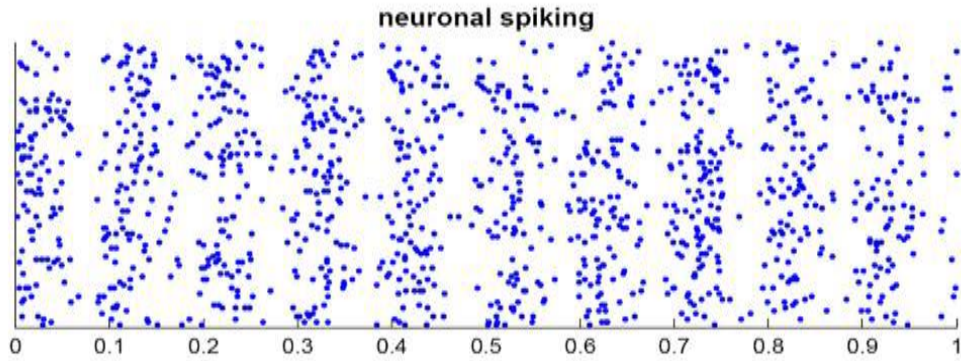
Taxonomy of Gray Matter Neurons

Xiaolong Jiang and Andreas Tolias at Baylor College of Medicine in Houston announced six new types of 15 adult mice brain cells by the method of slicing razor-thin slices of mature brain. (William Herkewitz, Science Nov 26, 2015) that establishing a complete census of all neuron cell types is of great importance in moving the field of neuroscience forward," says Tolias, at Baylor College of Medicine
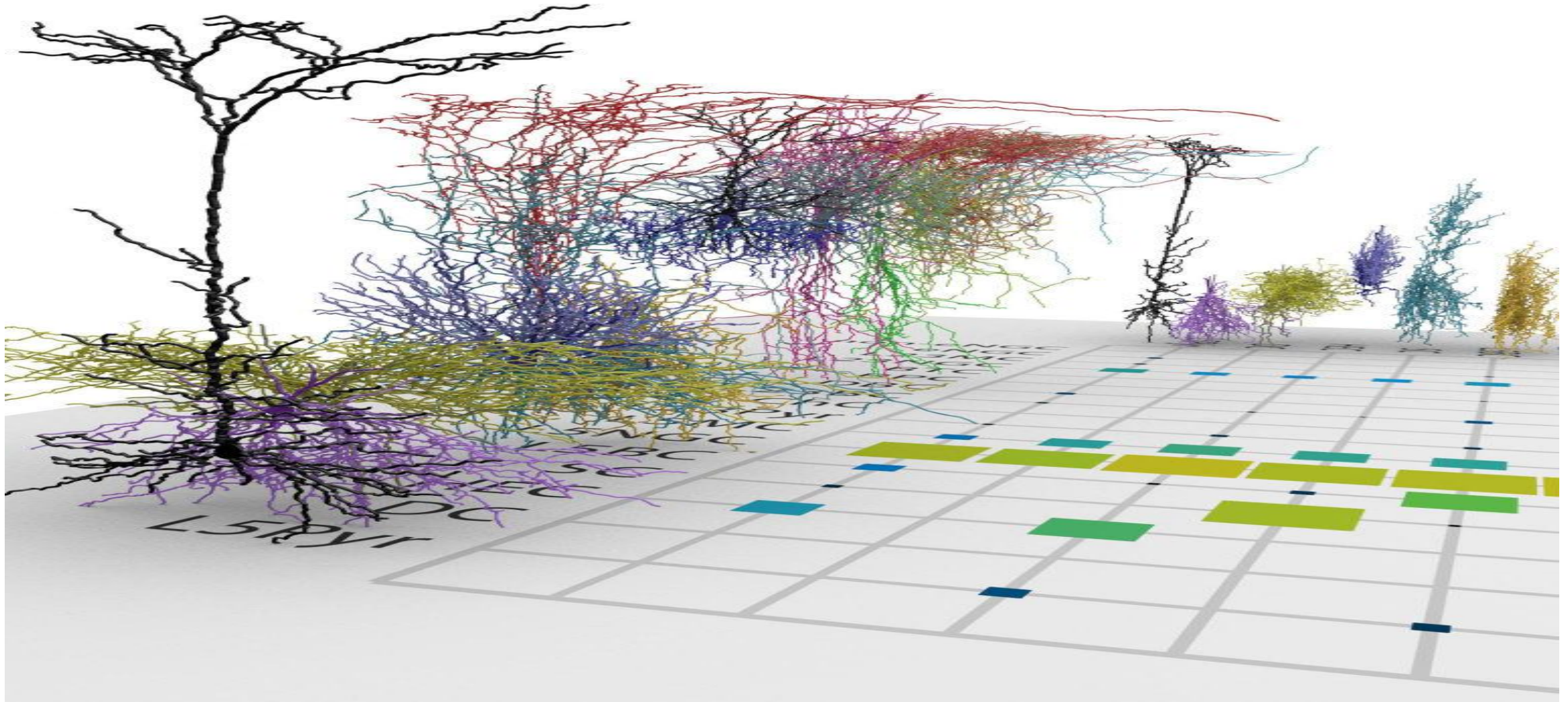
Biological Neural Networks (BNN) takes two to Tangle (neuro, glial cells). **Natural Intelligence (NI) is based on BNN**; Every neuron has 2 to 4 silent servant glial cells. Both are cell defined as
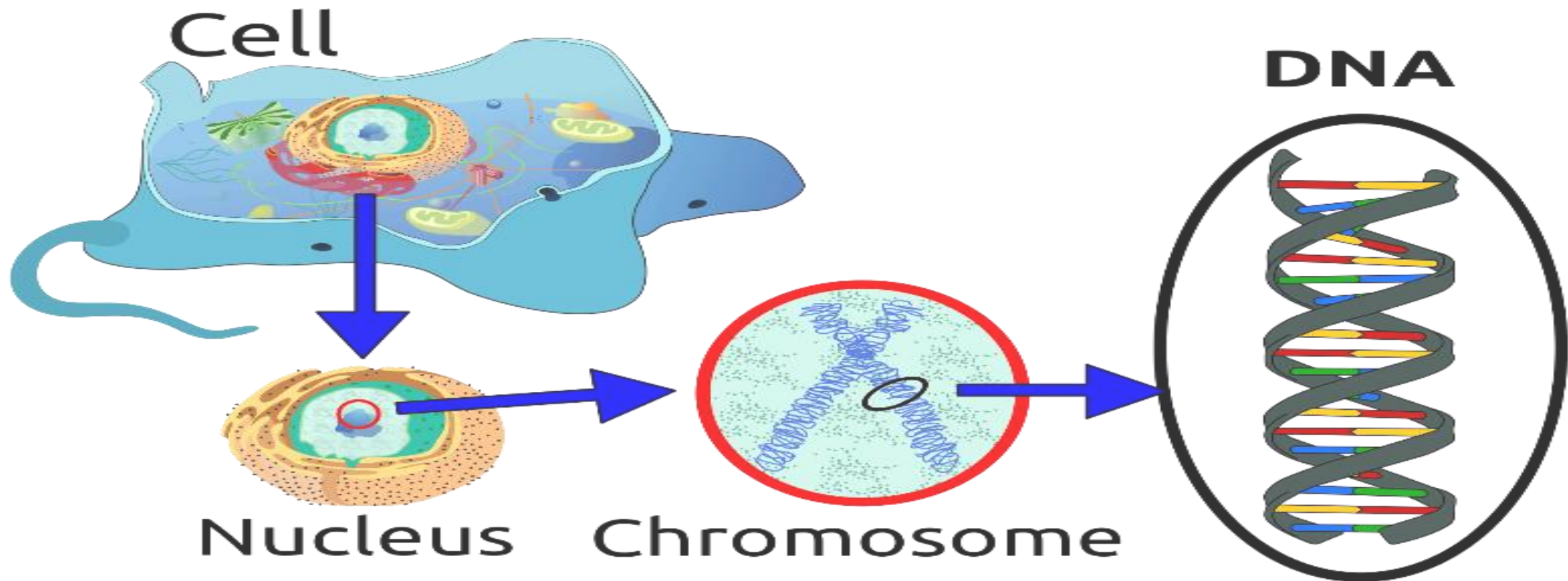
Mattson Restrictive Calorie; Blackburn Telomere Length; Yamanaka Induced Pluripotent Stem cells

3 billion pairs of A-T C-G codons of DNA packed in 23 pairs in a total of 46 chromosomes un-winded linearly in 3m long

2 mm to 20 mm
X-X female, X-Y male

Kenyon of UCSF discovered worm DAF-2 aging gene; DAF-16 for Longevity gene ;
but we're complex

We lose telomere size i.e. 50 times in 9 months known as Hayflick limit

What We Lose With Age

Chromosome

As cells divide over time...

Cell

Telomere, a protective covering

...telomeres shorten, and eventually cell division stops.

Source: WSJ release
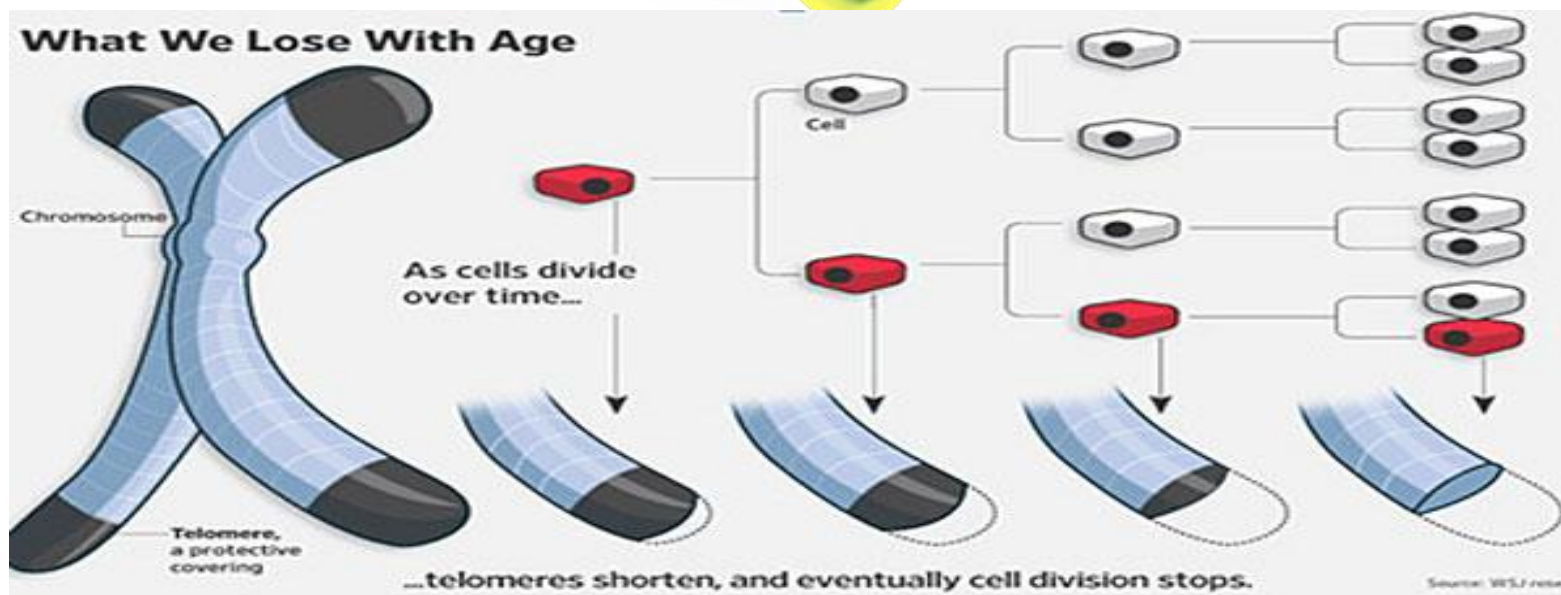
Eukaryote ("Eu=well" has a nuclei & Mitochondria foreign energy production cells inside our cell)
billions years ago; Prokaryote ("Pro=pre") primitive cell

TIME

FEBRUARY 23 / MARCH 2, 2015

SPECIAL **HEALTH** DOUBLE ISSUE

THIS BABY COULD LIVE TO BE **142 YEARS OLD**

Dispatches From the Frontiers of Longevity

time.com

13

Longevity attribute to Nonstop Production of Telomerase Enzyme; Mattson Calorie Restriction; Yamanaka induced Pluripotent Stem (iPS) 4 Genes. We anticipate Scotland Dolly the Sheep cloned at mother 6 years old somatic cells nuclear tfr, should be winding the clock back to embryonic state using iPS (TBD).

Cf. W. Duan, et. al. M. P. Mattson (Dir. NIA), "Dietary (Calorie) Restriction Normalizes Glucose Metabolism and Brain-Derived Neurotrophic Factor Levels, Slows Disease Progression and Increases Survival in Huntington Mutant Mice" PNAS Feb. 10 2003; Nobel 2012 Shinya Yamanaka 4 Genes induced Pluripotent Stem Cells of mice & human.

- **Lobsters (non-stop growth of telomerase keeping longevity) can live 150 yrs. On the contrary, Mammals have high growth rates in embryonic and juvenile phases and no growth in adult old phases.**

2012 Nobel Prize in Physiology or Medicine

Dolly the Sheep
A Brief History on Dolly
- In 1996, Dolly the Sheep was the first cloned animal
- She was cloned by the process called: somatic cell nuclear transfer (SCNT)
- Though having been only 6.5 years old before being euthanized, her legacy allowed many other large animals to be cloned.
- The idea of human cloning began with the success of Dolly the Sheep.
- She was euthanized due to a progressive lung disease called Sheep Pulmonary Adenomatosis (SPA)
- It is speculated that she could have been born (cloned) as a 6 year old

Shinya Yamanaka
University of Kyoto, Japan
Photo Credit:
Center for iPS cell Research and Application, Kyoto University

John B. Gurdon
Gurdon Institute in Cambridge, UK

"**Longevity of lobsters is linked to ubiquitous telomerase expression**," Klapper, et al. U. *Kiel*, FEBS Letters V. 439, pp.143–146 (1998)

.

Galápagos giant tortoise of **190 years**; **Arctic Clam of 400 years**

Le télomère protège le chromosome contre les dégradations

Elizabeth H. Blackburn   Carol W. Greider   Jack W. Szostak

Prix Nobel de physiologie ou médecine, 2009

Toward Personnel Medicine
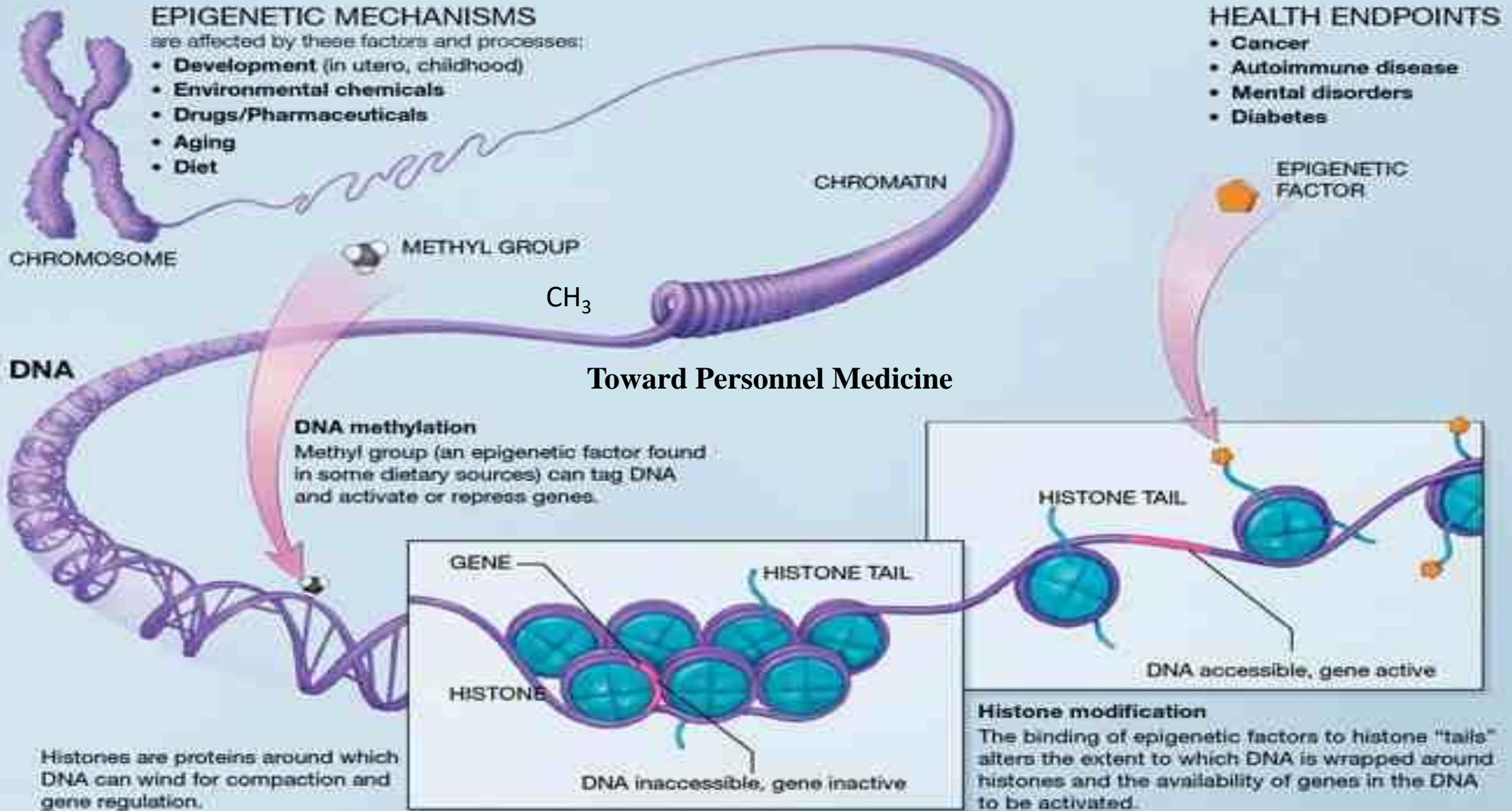
What is the **missing half of Einstein brain** ?

Einstein brain has 10 billions neurons 3 lb weight, like an ordinary human, but what made him smarter because of 100 billions Glial cells. Each glial is about 0.1 size of a neuron; electrical-silent, diligently housekeeping as neuron output servant cells (NI involves glial cells, as proved first by Szu at Min. Free Energy (MFE) from Boltzmann heat death $\because \Delta S_{entropy} > 0; T_o = 37^o_C = 310^o_K = constant$

$$\therefore \Delta H_{brain} \equiv \Delta E_{brain} - T_o \Delta S_{entropy} \leq 0;$$

$\because$ Hebb rule; $\therefore \vec{g} \equiv -\frac{\Delta H_{brain}}{\Delta \vec{D}endrite} \geq 0$ for 6 kinds brain cell functions has 6 types of Dendrites and

6 kinds glial ; The ratio generates 6 kinds dynamics
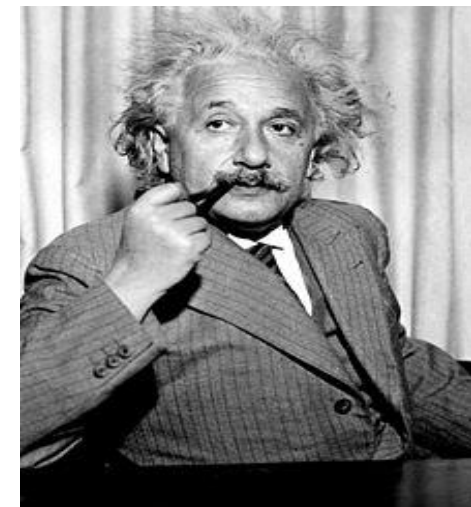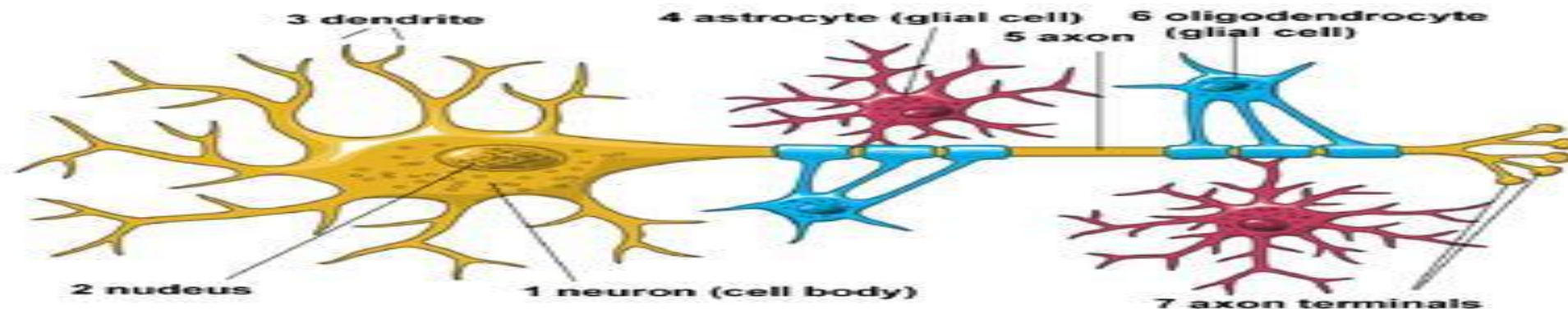
4=2 Glial cells:
Missing half of Einstein brain are Major Six Glial (Greek: Glue) Cells.
(4 CNS: Ependymal, Oligodendrocytes, Astrocytes, Microglia; 2 PNS: Satellite, Schwann cells)

Types of Neuroglia

Central Nervous System | Peripheral Nervous System
- Ependymal cells
- Oligodendrocytes
- Astrocytes
- Microglia
- Satellite cells
- Schwann cells
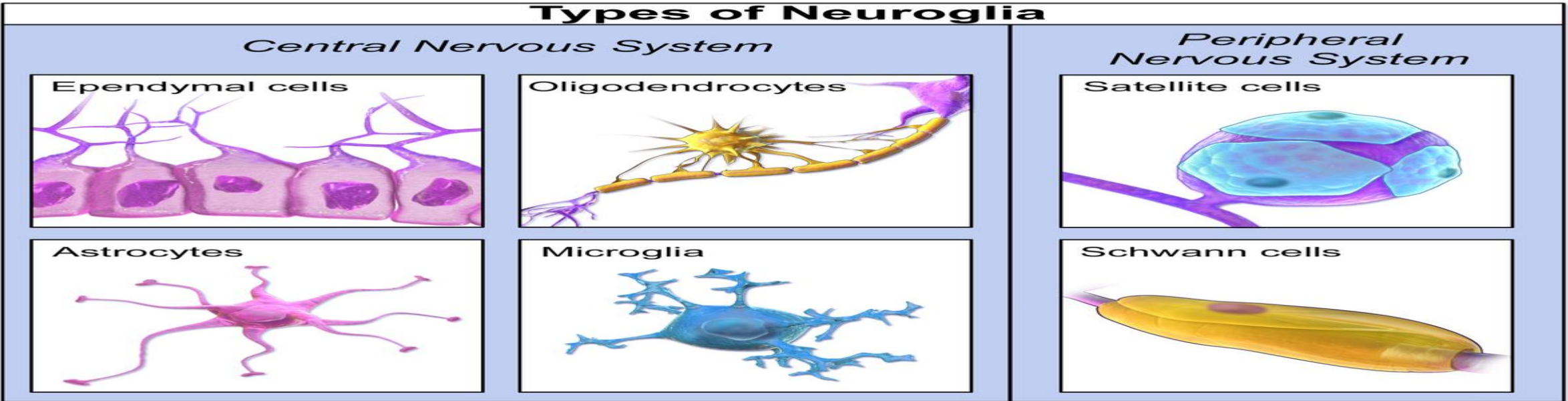
This was around same time that cell counts in the brain revealed glial cells to be nearly 90% of the brain including the white matter my sheath (we only use 10% of our brain comes from communication mediated by electromagnetic pulses. **Santiago Ramon Cajal**, Neuron Doctrine received the **Nobel Prize 1906**, while his brother **Pedro Cajal** supported Astrocytes type of **glial cells** supporting neuro transmitters that can ignite astrocyte calcium release.

Functionality of Glial Cells insures real time communication

(1) **Real Time (RT) Communication is possible** because Axon ion vesicles are confined and aligned up in **Axon Cable surrounded by electrically insulated Glial Cells** known as **White Matter Myelin Sheath**, becoming **as if a solid chain "one pops in one pops out" to increase the RT propagation speedway**, $O(\Delta t)$ =mille sec. for the survival of the species.

(2) Single layer Input $\left[W(\epsilon t)_{i,j}\right]\vec{X}_{j,in}(t) = \vec{S'}_{i,out}(t)$ *Multiple layer output* $\vec{S'}_{i,out}(t) \Longrightarrow Dendrite_i = \sum_j \left[W(\epsilon t)_{i,j}\right]\vec{S}_{j,out}(t),$

(3) 4+2 kinds Glial, glue, cells composed of white matter, Myelin sheath, making axon insulator an co-axial cable

Theoy of Natural Intelligence (NI),

iff 1. **Homeostasis** at Minimum Free (Helmholtz)Energy (MFE) $H_{Brain} = E_{Brain} - T_o S_{Brain}$ (Homosapiens 37^0C > Chicken 40^0C)

2. **Power of Pairs** $\vec{X}(t)$ irreversible dynamics "**Agree, Signal; Disagree, Noise;**" The Internal Knowledge Representation (IKR) is obtained naturally as the Boltzmann Entropy S as a measure of the **degree of uniformity of firing rates**, always increases.
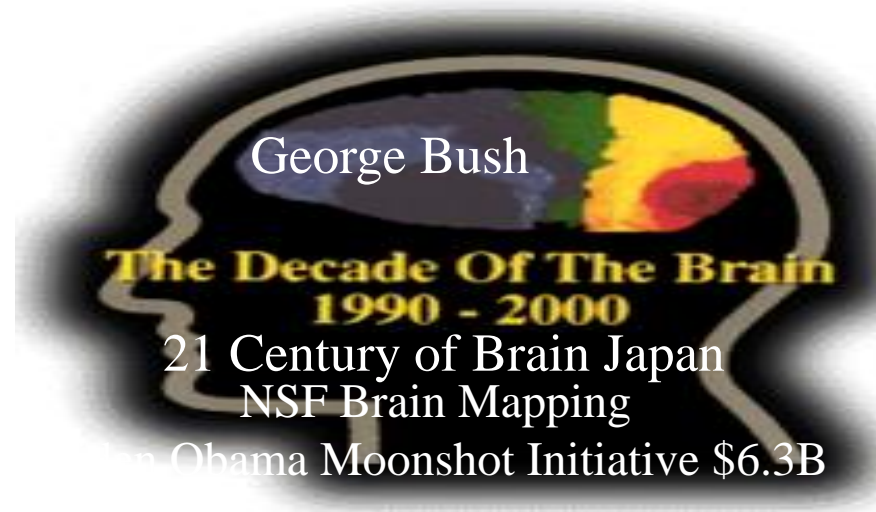
5 sensor pairs co-incidence account: "Agree, Signal; Disagree, Noise;" ABSTREACTION

George Bush

The Decade Of The Brain 1990 - 2000

21 Century of Brain Japan
NSF Brain Mapping
Obama Moonshot Initiative $6.3B

Cortex 17 (V1~V6 deep layers learning for Feature Extraction (color, edge, contour, texture, etc.) followed with Hippocampus Associative Memory recalls GENERALIZATION

Biological Neural Networks (BNN) of Homosapiens consisting of 10th billions of neurons & 100th billions of glial cells building the Natural Intelligence(NI) for unsupervised deep learning (UDL) spontaneity & de-confliction by elimination to focus on survival options

Paradox: Arrow of Time Irreversible Thermodynamics

(a) Ludwig Boltzmann; $S_{tot} = k_B \, Log \, W_{tot}$; circa 1890 for the irreversible thermodynamics $\Delta \vec{S} > 0$ heat death

(b) Henri Poincare the dynamics is time reversible , $m_o \dfrac{d^2 \vec{X}}{dt^2} = m_o \dfrac{d^2 \vec{X}}{d(-t)^2}$; $\pm i\hbar \dfrac{\partial \Psi}{\partial (\pm t)} = -\dfrac{\hbar^2}{2m} \nabla^2 \Psi$

(c) We now know Boltzmann is right, the trajectory is more than dynamics but initial boundary conditions which are irreproducible.

Quiz: What is the entropy? Which has a larger entropy? Sands or Rocks? Boltzmann said the Entropy is a measure of the degree of uniformity.

Let's consider a beautiful beach White Sands (e.g. Australia) versus Mountain Top Rocks, assuming that both composed of similar atoms silicon oxide and impurities.

Answer: Beach white sand has obviously more degree of uniformity, versus the original before eroded at mountain top rocks that have stronger variation of molecular *binding energy* known *archeologically* to the *paleontologist* as the information. Since Neal's third law of thermodynamics: Kelvin temperature can never be absolute zero, incessant molecular collisions that will mix toward maximum uniformity, irreversible increase of the entropy. In other words, the inter-molecular collision will decay gradually the molecular binging energy, namely the archeology information dear to paleontologist at hearts.

Derivation from Boltzmann Head Stone measure of the degree of uniformity & to his assertion of heat death

$$Boltzmann\ S_{tot} = k_B\ Log\ W_{tot} = S_{resovoir} + S_{brain};$$

$$W_{tot} \equiv exp\left(\frac{S_{tot}}{k_B}\right) = exp\left(\frac{S_{resovoir}T_o + S_{brain}T_o}{k_BT_o}\right) = exp\left(\frac{-E_{brain} + S_{brain}T_o}{k_BT_o}\right)$$

$$\equiv exp\left(\frac{-H_{brain}}{k_BT_o}\right) \equiv Maxwell\ Boltzmann\ Canonical\ Probability$$

$$k_BT_o \approx \left(\frac{1}{50}\right)eV$$

Use is made of the energy *conservation* : $\Delta E_{brain} + T_o\Delta S_{resovoir} = 0$

Helmholtz free to do energy responsible for the brain wave thinking work

$H_{brain} \equiv E_{brain} - T_oS_{brain}$ **similar to gasoline piston engine with exhaust**

$\Delta S_{brain} > 0$ **heat death without replenishment, Thus the basic learning principle is**

$$\Delta H_{brain} \leq 0$$

Learn experience changing synaptic impedance

$$\vec{S}_{brain\ output} = [W_{i,j}]\overrightarrow{X_{in}}(t);\ \textbf{hideen layer input}\ \vec{D}entritic_j \equiv \sum_i[W_{j,i}]\vec{S}'_i;$$

Unsupervised Learning Cost Function MFE

$$\Delta H_{brain} \equiv \Delta E_{brain} - T_o \Delta S_{brain} \leq 0$$
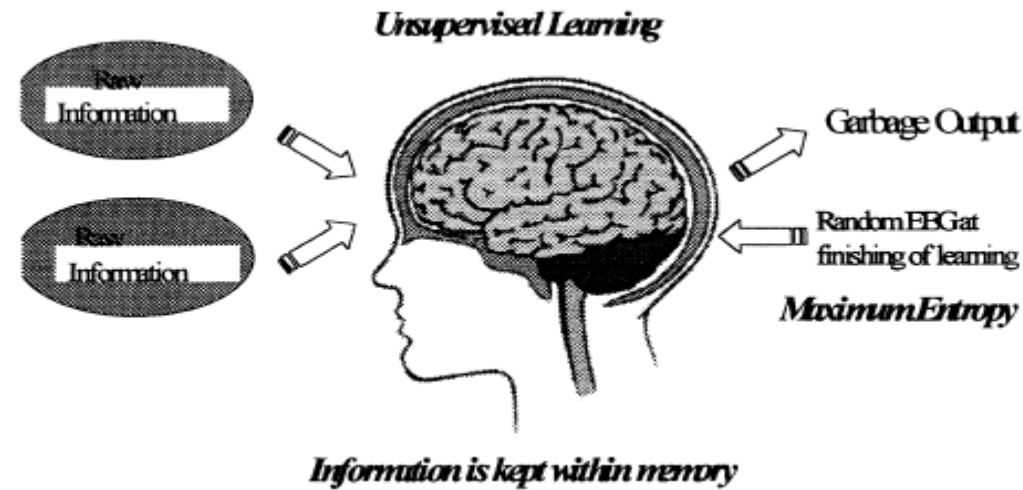
Boltzmann Equilibrium Thermodynamics

$$S_{tot} \equiv k_B \, Log \, W_{tot}; W_{tot} \equiv \exp\left(\frac{S_{tot}}{k_B}\right) = \exp\left(\frac{(S_{brain} + S_{env.})T_o}{k_B T_o}\right) = \exp\left(\frac{S_{brain}T_o - E_{brain}}{k_B T_o}\right)$$

$$= \exp\left(-\frac{H_{brain}}{k_B T_o}\right) \equiv W_{MB}$$

Maxwell Boltzmann Canonical Probability is the statistical equilibrium is known as Homeostasis

$$H_{brain} \equiv E_{brain} - T_o S_{brain}$$

$$\Delta H_{brain} \equiv \Delta E_{brain} - T_o \Delta S_{brain}$$

Boltzmann heat death $\Delta S_{brain} > 0$ due to incessant collisional mixing towards more uniformity (Inequilibrium Thermodynamics)

$$\Delta H_{brain} \equiv \Delta E_{brain} - T_o \Delta S_{brain} \leq 0$$

satisfied the MFE irreversible
thermodynamics $\Delta H_{brain} \leq 0$ in the
Lyaponov convergence control theorem
sense, whether the Synapses memory
~~gains or losses~~ $\Delta[W_{i,j}] \leq 0$:

NEURON

Dendrites (receivers)

Axon Terminals (transmitters)

Schwann's Cells (they make the myelin)

Node of Ranvier

Cell Body

Nucleus

Axon (the conducting fiber)

Myelin Sheath (insulating fatty layer that speeds transmission)

©EnchantedLearning.com
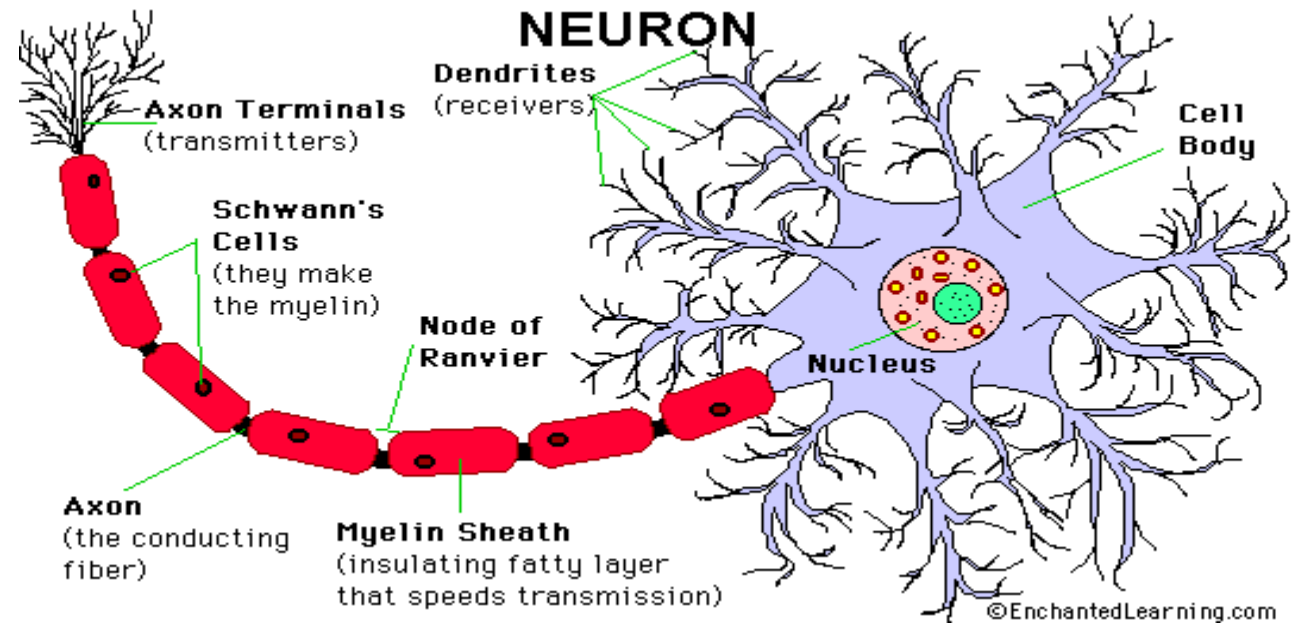
$$\frac{\Delta H_{brain}}{\Delta t} = \left(\frac{\Delta H_{brain}}{\Delta[W_{i,j}]}\right)\frac{\Delta[W_{i,j}]}{\Delta t} = -\frac{\Delta[W_{i,j}]}{\Delta t}\frac{\Delta[W_{i,j}]}{\Delta t} = -\left(\frac{\Delta[W_{i,j}]}{\Delta t}\right)^2 \leq 0$$

If and only if Newtonian acceleration/deceleration is driven by the MFE Gradient Force is
satisfied by the Biological Neural Network (BNN) Dynamics

$$\frac{\Delta[W_{i,j}]}{\Delta t} = -\frac{\Delta H_{brain}}{\Delta[W_{i,j}]} \lesseqgtr 0$$

25

Canadian Biologist Donald. O. Hebb Learning Rule :1904-1985

26

Dendrite net input vector is morphologically dependent of the neuron cell typing

$$\overrightarrow{Dendrite_j} \equiv \sum_i [W_{i,j}]\vec{S}'_i$$

$$\frac{\Delta[W_{i,j}]}{\Delta t} = \left(-\frac{\Delta H_{brain}}{\Delta \overrightarrow{Dendrite_j}}\right)\frac{\Delta \overrightarrow{Dendrite_j}}{\Delta[W_{i,j}]} \approx \vec{g}_j\vec{S}'_i,$$

$$\vec{g}_j \equiv -\frac{\Delta H_{brain}}{\Delta \overrightarrow{Dendrite_j}}$$

We have derived for the first time the mathematical definition of glial cell (discovered by pathologist **Rudolf Virchow 1856).** This math definition of glial cells (Szu 2017 US PTO) explains the unified theory of all different neuron typing about 15 in mice, morphology predicts different Dendrite vectors defining different kinds of Glial Cells, led to all kinds of Glial cells (4 in CNS: Ependymal, Oligodendrocytes, Astrocytes, Microglia; 2 in PNS: Satellite, Schwann cells).The **divergence of glial may explain the brain tumor Glioma!**. Consequently, from BNN we obtain Hebb learning update rule($\eta \approx O(\Delta t)$) for a single layer of BNN.

$$\Delta[W_{i,j}] = [W_{i,j}(t+1)] - [W_{i,j}(t)] = \vec{g}_j\vec{S}'_i\eta$$

Rosette Stone 2016
Math Theory of Deep Learning in BNN NI or ANN AI

$$\text{NI glial } g_i \equiv -\frac{\partial H}{\partial Dendrite_i} \qquad \text{AI delta } \delta_i \equiv -\frac{\partial LMS}{\partial net_i}$$

Einstein at Princeto

$$Dendrite_j \equiv \sum_i [W_{j,i}]S_i; \qquad net_j \equiv \sum_i [W_{j,i}]O_i;$$

$$S_i = \sigma(Dendrite) \qquad O_i = \sigma(net_i)$$

3 dendrite   4 astrocyte (glial cell)   6 oligodendrocyte
5 axon   (glial cell)

2 nucleus   1 neuron (cell body)   7 axon terminals
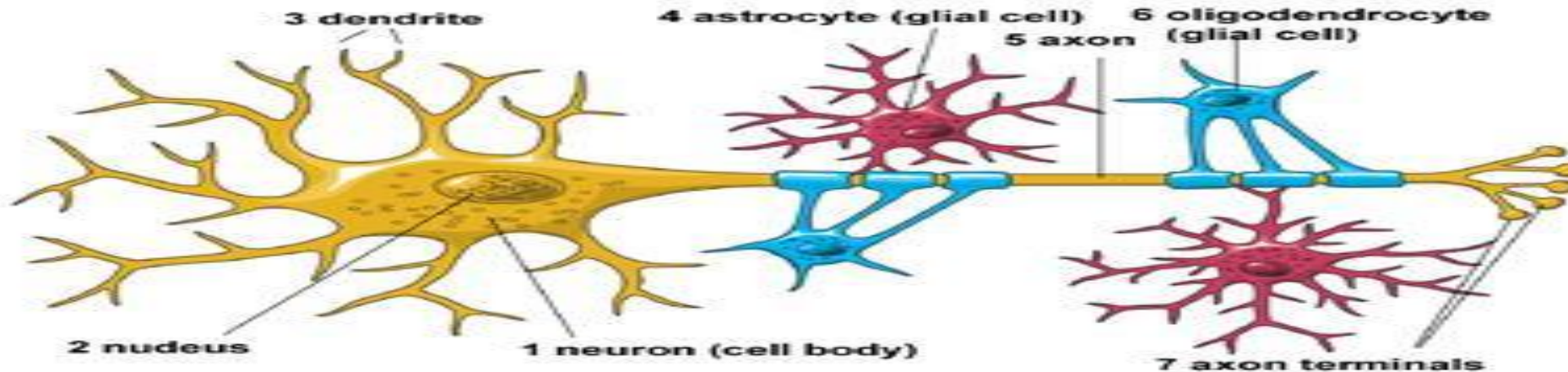
Supervised & Unsupervised Deep Learning

What is the Cost Functions for supervised and unsupervised DL ?

- **Supervised DL** utilizes **the Least Means Squares (LMS)** errors for *Artificial Intelligence (AI), ANN learnable relational databases*;

- **Unsupervised DL** utilizes *the Minimum Free Energy (MFE)* at **BNN** at Helmholtz **MFE** for *Natural Intelligence (NI), if and only if*

- *(i) Isothermal Brain (homosapiens at 37^0C (optimum for hemoglobin elasticity);* while Chicken 40^oC *(for egg hatching); but lacking of big thumb for holding* tools becomes less intelligent than homosapiens *(we ate them, not vice versa, Q.E.D.).*

- *(ii) Power of Pairs (eyes, ears, nostrils, taste buds, tactile pre-processing "agree, signal; disagree, noise"): BNN Learning* $[W_{i,j}]\vec{X}_{in,pair}(t) = \vec{S}_{out,fusion}(t)$

AI LMS Error cannot separate figure from the ground

NI Human Target Recognition must be able separate binary Figure and Ground under the dusk dim light far away. This could be any simple Ambiguity Figures for computational simplicity. The idea of NI in BNN for the survival is manifested clear in "$\vec{F}$ =Tigress" & Ground "$\vec{G}$ =Tree". In contrast, the supervised cost function is LMS AI based on ANN becomes ambiguity of binary figure & ground Least Mean Squares (LMS) cost function $|(\vec{F} - \vec{G})^2|=|(\vec{G} - \vec{F})^2|$ could not separate to run away for the survival of the species due to the switch of the algebra sign. However, higher order of moment expansion of MFE can separate the tiger and tree in the remote dime light for the survival of Homosapiens.

The linear term can already tell the difference between the target lion versus the background tree, $(0 - 1) \neq (1 - 0)$ without suffering the LMS parity invariance: $(0 - 1)^2 = (1 - 0)^2$

Biological Neural Net (BNN), weighted dendritic sum (bold arrow vector input to a neuron $Dendrite_i = \sum_j [W(\epsilon t)_{i,j}] \vec{S}_{j,out}(t))$ and electrically neural Glial Cells housekeeping servant cells supported real-time co-axial Axon cable output e.g. HVS Cortex 17 has layers V1~V4
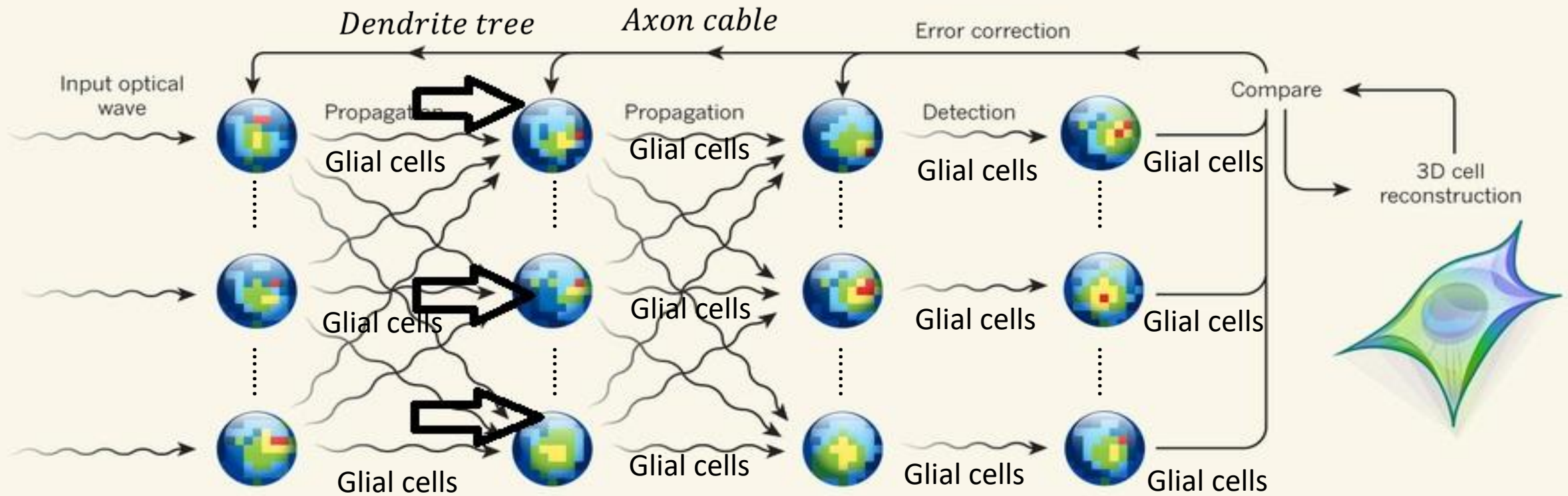
Analytic Proof of Glial Cells
Taylor Expansion defines the negative slope as the glia cells due to Biologist D.O. Hebb learning rule circa 1950

$$H_{brain} = H_o\left(\vec{X}(t_o)\right) + \sum_{i=1}^{N'} (\frac{\partial H_{brain}}{\partial S_i})_o\{ S_i - S_i(t_o)\} + \sum_{j}^{N'}\sum_{i=1}^{N'} (\frac{\partial^2 H_{brain}}{\partial S_j \partial S_i})_o\{ S_i - S_i(t_o)\}\{ S_j - S_j(t_o)\} + O(|\Delta\vec{S}|^3)$$

$$H_{brain} = H_o + \vec{g}_{glial}\{[W]\vec{X}_o - \vec{S}\} + \frac{\partial \vec{g}_{glial}}{\partial \vec{S}} * \{[W]\vec{X}_o - \vec{S}\}\{[W]\vec{X}_o - \vec{S}\}$$

- **D.O. Hebb observed bi-linear learning rule (dendritic Input glial, g_j, -MFE slope, output Axon firing rate S_i MFE itself):**

$$\frac{\partial [W_{ji}]}{\partial t} = -\frac{\partial H}{\partial [W_{ji}]} = -\frac{\partial H}{\partial Dendrite_j}\frac{\partial Dendrite_j}{\partial [W_{ji}]} = = g_j\, S'_i;\ where\ bi\text{-}linear\ Hebb\ rule$$

follows:

$$Input\ g_{glialj} = -\frac{\partial H}{\partial Dendrite_j};\ output\quad \frac{\partial Dendrite_j}{\partial [W_{ji}]} = \frac{\partial}{\partial [W_{ji}]}\Sigma_i[W_{j,i}]S'_i = S'_i.$$

Sigmoid Logic & Exact solution of Control Ricatti Equation

Single layer Input $\left[W(\epsilon t)_{i,j}\right]\vec{X}_{j,in}(t) = \vec{S}'_{i,out}(t)$

Multiple layer output $\vec{S}'_{i,out}(t) \Rightarrow Dendrite_i = \sum_j \left[W(\epsilon t)_{i,j}\right]\vec{S}_{j,out}(t)$

Sigmoid Logic $\sigma(H^*_{brain} - H_{brain}) = \dfrac{exp\left(\frac{-H^*_{brain}}{k_B T_o}\right)}{exp\left(\frac{-H^*_{brain}}{k_B T_o}\right) + exp\left(\frac{-H_{brain}}{k_B T_o}\right)} = \dfrac{1}{1 + exp\left(\frac{H^*_{brain} - H_{brain}}{k_B T_o}\right)}$

$H^*_{brain} - H_{brain} = Dendrite_i = \sum_j \left[W(\epsilon t)_{i,j}\right]\vec{S}_{j,out}(t)$

$\sigma(Dendrite_j) = \vec{S}'_j;$

$\frac{dy}{dx} = \frac{d}{dx}(1 + exp(x))^{-1} = -1((1 + exp(x))^{-2}([exp(x) + 1] - 1) = y^2 - y = y(y - 1);$

$$\sigma'_j(Dendrite_j) = \vec{S}'_j(\vec{S}'_j - 1)$$

is a limiting window function [0, 1] 1 for 100 Hz

Back Prop of MFE slopes glial cells

$$g_j \equiv -\frac{\partial H}{\partial Dendrite_j} = -\frac{\partial H}{\partial S_j}\frac{\partial S_j}{\partial Dendrite_j} = -\frac{\partial H}{\partial S_j}\sigma_j'(Dendrite_j)$$

$$-\frac{\partial H}{\partial S_j} = -\sum_k \frac{\partial H}{\partial Dendrite_k}\frac{\partial Dendrite_k}{\partial S_j}$$

$$= -\sum_k \frac{\partial H}{\partial Dendrite_k}\frac{\partial}{\partial S_j}\sum_i [W_{k,i}]S_i = -\sum_k \frac{\partial H}{\partial net_k}[W_{k,j}] = \sum_k g_k[W_{k,j}]$$

$$g_j = \sigma_j'(net_j)\sum_k g_k[W_{k,j}]$$

Q.E.D.

$$\text{MFE } \Delta H_{brain} \equiv \Delta E_{brain} - T_o \Delta S_{brain} \; ; \& \text{ Irreversible } \Delta S_{brain} > 0$$

**BNN, Natural Intelligence (NI) & Deep Learning (Multiple Layers)
Back Prop Learning Rule:**

$$\begin{aligned} \left[W_{ji}(t+1) \right] \\ = \left[W_{ji}(t) \right] + \eta \, \vec{g}_j \vec{S}_i' + \alpha_{momtum} \left[W_{ji}(t) - \left[W_{ji}(t-1) \right] \right] \end{aligned}$$

Q.E.D.

where ad hoc momentum term is to bypass the Mexican standoff local minimum

Future Perspective of Intelligent Machine
New Consumer Index Machines IQ= AI/NI

1. MIQ=10% is loyal to its human master and its own survivability to differentiate electric power plug having a two-porn's of 110 Volts or three porn's of 220 Volts.2.

2. MIQ=20% is able to understanding human conversation in a fixed semantic network for a closed domain dialogue.

3.**MIQ=30%** is able to read facial expression and voice tone for e-IQ to understanding the emotion need of human being.

4. MIQ=40% is able to command and control a small team of other robots.

5. MIQ=50% is able to "explore the tolerance of imprecision," e.g. using fuzzy logic to negotiate a single precision path finding in an open save terrain.

BDA: drug discovery with the help of system biology & UDL

Is the Herbal Mushroom G Lucidum, Lingzhi (that 2000 Nobel Laureate Literature Mr. Gao Xingjian recovered in cancer) similar to Merck immunotherapy Keytruda (Pembrolizumab) drug (that President Jimmy Carter Liver and Brain Metastasis cancer: Aug. 2015 ~Feb. 2016)?

While Merck drug (Yellow balls) are targeted at the Programmed cell Death 1 (PD-1) receptor and allows the body's own immune system go after the cancer cells. While they are all worked on human immune systems, the key difference between Oriental Herbal Medicine and Western Molecular personalized precision targeted drug is mainly in that the holistic is slow in nature of herbal drug for years versus fast drug in half a year..

cancer

Photo©Raija Tuomainen
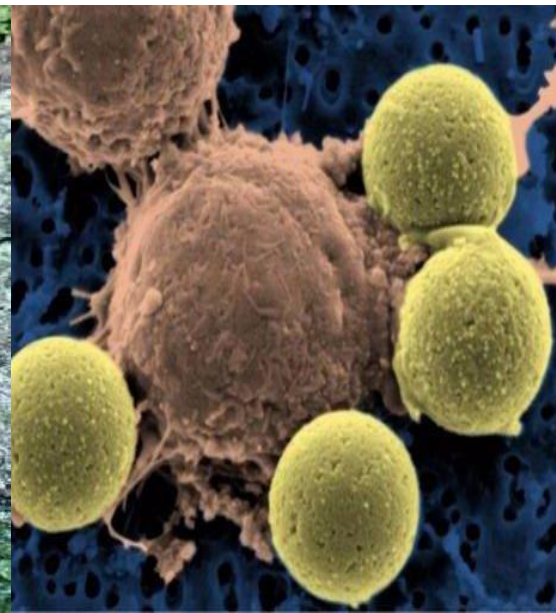
Looking forward we need for BDA drug discovery, law enforcement both Deep Learning for Feature Extraction & Machine Learning transform salient feature to Support Vector Machine for Classifier

- We keep i-Robot without e-IQ loving feeling toward master & jeopardies spouse. Then, no Terminator III danger of AI control the world. In general, $MIQ \subseteq 50\%$ should be in a supervised learning category using large data basis training by the lookup table setup. Since no e-IQ in MIQ.

- Machine intelligence about 20% to 30% should understand simple language & recognize human emotion and some extrapolation and interpolation capability.

- Other than those robots working on the factory floor, the futurist intelligent robots happen in **Deep Ocean, Outer space, Melt down Reactor** with an **unforeseeable NL dynamics interwoven with non-stationary complexity**.

- They may need $MIQ \supseteq 50\%$ When $MIQ \supseteq 50\%$ a machine shall behave human-like with some intrinsic *e-IQ* for better machine-human interface, machine sensory are equipped with learning without the supervision. For example, the team **MIQ** of *UXV (X=air, ground, marine)* shall be equipped with the *Swarming Intelligence* for unsupervised self-organization team formation.

Appendix A Modern Brain Physiology Biology:

(i)   Biological Neural Net based on Thermodynamics Brain at Minimum Free Energy (MFE);

(ii)   Smartness inheritance? Genome by DNA & Phenome by Epigenetic;

(iii) Heathy Longevity by Telomerase Enzyme.

(iv)   Brain Tumor is known as Glioma. A **glioma** is a type of **tumor** that starts in the **brain** or spine. It is called a **glioma** because it arises from glial cells. The most common site of **gliomas** is the **brain**. **Gliomas** make up about 30% of all **brain** and central nervous system **tumors** and 80% of all malignant **brain tumors**. (https://en.wikipedia.org/wiki/**Glioma**)

Appendix B "Science has nothing to do with the truth, but the consistency," Albert Einstein circa 1910
we can unified Sources of Attractive Field Theory: electron radius, gravitational diameter, and glial cell size.

$$\frac{ee}{r_o} = E = m_e C_o^2; \quad r_o = \frac{e^2}{m_e C_o^2} = 2.8 \, 10^{-13} \text{cm}$$

$$F = -m_o g = -G \frac{m_o M}{|d_o|^2}; \quad E = m_o C_o^2; \qquad d_o = \frac{1}{C_o}\sqrt{\frac{GME}{F}} = \frac{2F}{\pi G M <\rho_o>}$$

$$\vec{g}_j \equiv -\frac{\partial H}{\overrightarrow{\partial Dentritic_j}} = O\left(\frac{neuron}{10 \; glials}\right)$$

$$\left|\overrightarrow{Dendritic_j}\right| \equiv \left|\sum_{i=1}^{N}[W_{i,j}] \vec{S}_i^{'(t)}\right| = size \; of \; glial = 0.1 \; \text{size of neuron; the size of}$$

H is the diameter of neurons in sub-millimeters. Inversely, we determine from observation of 0.1 neuron size, the Glue force is 0.1 chemical affinity of neuron layers

Appendix C Whether the smartness of NI is endowed or cultured?
Epigenetic Methylation (Phenome) versus Genetic (**A-T C-G codons** Genome)?

- Older & Wiser (Longevity gain more experience & more judgment.)
- Is NI coded in Genome DNA A-C, G-T pairs or Phenome Methylation (**3 billion pairs of A-T C-G codons of DNA packed in 23 pairs in 46 Chromosomes un-winded linearly in 3m long**)
- European (EU) **Human Epigenetic Programs (HEP)** have investigated many identical twins that their lifestyles may have influenced the epigenetics that pass down to influence the next gen.
- The United Sates (US) **Human Genome Program (HGP)** has decoded the full human genome.
- If Genetic DNA like a **hardware**; then Epigenetic is like a **software**. We need both the US/HGP & the EU/HEP.

Appendix D: Morphological Learning

(1) Beer Belly architecture for a large internal representation degree of freedom (d. o. f.) for easy Generalization of numerous representations capability in compression to the output class.

(2) Hourly glass architecture with less d. o. f. for Abstraction

(3) Input data-driven determines connectivity morphology by glial cells two ad hoc principles

(i)   Use it or Lose it, Pruning Node or Growth Connect Node may be decided by difference in MFE's with it or without

(ii) Hot spot traffic jam recruit more neurons

Appendix E: Nothing but living style can beat a good gene for health longevity: Error Correction in walking meditation can keep the Telomerase enzyme in dynamic balance of telomeres to prevent a premature aging &/or carcinoma.

- In contrary to popular belief, the hydrogen bonds do not stabilize the DNA, and stabilization is mainly due to 3-D stacking (*epigenetic phenome* helps the **DNA genome**). The **back-chaining proofreading** can reduce error rate below Maxwell-Boltzmann Probability (MBP), 1 error in 10,000 ten thousands

$$MBP_{CG} = \exp{-\left(\frac{\Delta H_{CG}}{k_B T}\right)} \leq MBP_{AT} = \exp{-\left(\frac{\Delta H_{AT}}{k_B T}\right)}$$

ΔH_{CG} is slightly higher than ΔH_{AT}, thus MBP_{AT} is more stable than MBP_{CG}

- BSS $\vec{X} = [A]\vec{S}$; $\vec{S} = [W]\vec{X}$. *e.g.* Given Beethoven first 3 notes: " 5, 5, 1...." :X=5= (0+5; 1+4; 2+3; 3+2; 4+1; 5+0) in MBP unit at $K_B T=1/40eV$ for T=300°; Find hidden sources tones 2=3 and 3+2 occurring twice that have the highest MBP 2 exp(-2/K_BT)exp(-3/K_BT) after ruling out rare *high energy* cases: 0+5 and 1+4, in favor of lower energy: twice 2+3.

- Given multiple spectral bands, MFE to find hot spot cancer or not by 1-layer Lagrange Constraint Neural Net (LCNN), where Lagrange coeff. are Glial cells ("Unsup. Learning at MFE," Szu, Miao, Qi SPIE Vol. 6576, 657605, (2007).

More on Hopfield back-chaining kinetic proofreading

for $DNA\ Codec\ \ P = \exp(-H/k_BT_o)$ *John Hopfield* *"Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity"*. *Proc. Natl. Acad. Sci. U.S.A.* **71** (10): 4135–9. Oct 1974

- Kinetic proofreading allows enzymes to discriminate between two possible reaction pathways leading to correct or incorrect products with an accuracy higher than what one would predict based on the difference in the activation energy between these two pathways.
- Similar DNA reading by tRNA of A-T & C-G pairs the energy difference is minutia. Backward chaining help increase the success reading.
- To achieve an error rate of e^{-10} requires several comparison steps. Hopfield predicted on the basis of this theory that there is a multistage ratchet in the ribosome which tests the match several times before incorporating the next amino acid into the protein.

Appendix F: Deep Learning Machine Learning Tensor flow Python Language or Math Lab by Math Works:
Dr. Joanna Pingel, "Object Recognition Deep Machine Learning for Computer Vision"

https://www.tensorflow.org/ Python Language for GPU

another video from MATHWORKS - szuharoldh@gmail.com - Gmail.html

Appendix B: Dr. Hesham M. Eraqi; hesham.eraqi@gmail.com Matlab Code (8 pages with comments)

http://heraqi.blogspot.com.eg/2015/11/mlp-neural-network-with-backpropagation.html

Appendix B: Multilayer Perceptron Feed Forward Fully Connected Neural Network with a Sigmoid activation function₁

```matlab
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Multilayer Perceptron (MLP) Neural Network Function using MATLAB:       %
%  An implementation for Multilayer Perceptron Feed Forward Fully        %
%  Connected Neural Network with a sigmoid activation function. The      %
%  training is done using the Backpropagation algorithm with options for  %
%  Resilient Gradient Descent, Momentum Backpropagation, and Learning    %
%  Rate Decrease. The training stops when the Mean Square Error (MSE)    %
%  reaches zero or a predefined maximum number of epochs is reached.     %
%                                                %
%  Four example data for training and testing are included with the     %
%  project. They are generated by SharkTime Sharky Neural Network       %
%   (http://sharktime.com/us_SharkyNeuralNetwork.html)            %
%                                                %
% Copyright (C) 9-2015 Hesham M. Eraqi. All rights reserved.          %
%             hesham.eraqi@gmail.com                 %
%                                                
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%% Clear Variables, Close Current Figures, and Create Results Directory
clc;
clear all;
close all;
mkdir('Results//'); %Directory for Storing Results
%% Configurations/Parameters
dataFileName = 'sharky.spirals.points'; %sharky.linear.points - sharky.circle.points - sharky.wave.points -
sharky.spirals.points
% nbrOfNeuronsInEachHiddenLayer = [10 10]; %linear:[4] - circle:[10] - wave,spirals:[10 10]
nbrOfNeuronsInEachHiddenLayer = [4 4]; %linear:[4] - circle:[10] - wave,spirals:[10 10]
nbrOfOutUnits = 2;
unipolarBipolarSelector = 0; %0 for Unipolar, -1 for Bipolar

learningRate = 0.15;
nbrOfEpochs_max = 500000;

enable_resilient_gradient_descent = 1; %1 for enable, 0 for disable
learningRate_plus = 1.2;
learningRate_negative = 0.5;
deltas_start = 0.9;
deltas_min = 10^-6;
deltas_max = 50;

enable_decrease_learningRate = 0; %1 for enable decreasing, 0 for disable
learningRate_decreaseValue = 0.0001;
min_learningRate = 0.05;

enable_learningRate_momentum = 0; %1 for enable, 0 for disable
momentum_alpha = 0.05;

draw_each_nbrOfEpochs = 100;

%% Read Data
importedData = importdata(dataFileName, '\t', 6);
Samples = importedData.data(:, 1:length(importedData.data(1,:))-1);
TargetClasses = importedData.data(:, length(importedData.data(1,:)));
TargetClasses = TargetClasses - min(TargetClasses);
ActualClasses = -1*ones(size(TargetClasses));

%% Calculate Number of Input and Output NodesActivations
nbrOfInputNodes = length(Samples(1,:)); %=Dimention of Any Input Samples
% nbrOfOutUnits = ceil(log2(length(unique(TargetClasses)))) + !; %Ceil(Log2( Number of Classes ))

nbrOfLayers = 2 + length(nbrOfNeuronsInEachHiddenLayer);
nbrOfNodesPerLayer = [nbrOfInputNodes nbrOfNeuronsInEachHiddenLayer nbrOfOutUnits];

%% Adding the Bias as Nodes with a fixed Activation of 1
nbrOfNodesPerLayer(1:end-1) = nbrOfNodesPerLayer(1:end-1) + 1;
Samples = [ones(length(Samples(:,1)),1) Samples];
```

http://heraqi.blogspot.com.eg/2015/11/mlp-neural-network-with-backpropagation.html

Multilayer Perceptron Feed Forward Fully Connected Neural Network with a Sigmoid activation function$_2$

```matlab
%% Calculate TargetOutputs %TODO needs to be general for any nbrOfOutUnits
TargetOutputs = zeros(length(TargetClasses), nbrOfOutUnits);
for i=1:length(TargetClasses)
    if (TargetClasses(i) == 1)
        TargetOutputs(i,:) = [1 unipolarBipolarSelector];
    else
        TargetOutputs(i,:) = [unipolarBipolarSelector 1];
    end
end

%% Initialize Random Wieghts Matrices
Weights = cell(1, nbrOfLayers); %Weights connecting bias nodes with previous layer are useless, but to
make code simpler and faster
Delta_Weights = cell(1, nbrOfLayers);
ResilientDeltas = Delta_Weights; % Needed in case that Resilient Gradient Descent is used
for i = 1:length(Weights)-1
    Weights{i} = 2*rand(nbrOfNodesPerLayer(i), nbrOfNodesPerLayer(i+1))-1; %RowIndex: From Node
Number, ColumnIndex: To Node Number
    Weights{i}(:,1) = 0; %Bias nodes weights with previous layer (Redundant step)
    Delta_Weights{i} = zeros(nbrOfNodesPerLayer(i), nbrOfNodesPerLayer(i+1));
    ResilientDeltas{i} = deltas_start*ones(nbrOfNodesPerLayer(i), nbrOfNodesPerLayer(i+1));
end
Weights{end} = ones(nbrOfNodesPerLayer(end), 1); %Virtual Weights for Output Nodes
Old_Delta_Weights_for_Momentum = Delta_Weights;
Old_Delta_Weights_for_Resilient = Delta_Weights;

NodesActivations = cell(1, nbrOfLayers);
for i = 1:length(NodesActivations)
    NodesActivations{i} = zeros(1, nbrOfNodesPerLayer(i));
end
NodesBackPropagatedErrors = NodesActivations; %Needed for Backpropagation Training Backward Pass

zeroRMSReached = 0;
nbrOfEpochs_done = 0;
```

```matlab
%% Iterating all the Data
MSE = -1 * ones(1,nbrOfEpochs_max);
for Epoch = 1:nbrOfEpochs_max

    for Sample = 1:length(Samples(:,1))
        %% Backpropagation Training
        %Forward Pass
        NodesActivations{1} = Samples(Sample,:);
        for Layer = 2:nbrOfLayers
            NodesActivations{Layer} = NodesActivations{Layer-1}*Weights{Layer-1};
            NodesActivations{Layer} = Activation_func(NodesActivations{Layer},
unipolarBipolarSelector);
            if (Layer ~= nbrOfLayers) %Because bias nodes don't have weights connected to
previous layer
                NodesActivations{Layer}(1) = 1;
            end
        end
        % Backward Pass Errors Storage
        % (As gradient of the bias nodes are zeros, they won't contribute to previous layer errors
nor delta_weights)
        NodesBackPropagatedErrors{nbrOfLayers} =  TargetOutputs(Sample,:)-
NodesActivations{nbrOfLayers};
        for Layer = nbrOfLayers-1:-1:1
            gradient = Activation_func_drev(NodesActivations{Layer+1}, unipolarBipolarSelector);
            for node=1:length(NodesBackPropagatedErrors{Layer}) % For all the Nodes in current
Layer
                NodesBackPropagatedErrors{Layer}(node) =  sum(
NodesBackPropagatedErrors{Layer+1} .* gradient .* Weights{Layer}(node,:) );
            end
        end
```

Multilayer Perceptron Feed Forward Fully Connected Neural Network with a Sigmoid activation function$_3$

```
% Backward Pass Delta Weights Calculation (Before multiplying by learningRate)
    for Layer = nbrOfLayers:-1:2
        derivative = Activation_func_drev(NodesActivations{Layer}, unipolarBipolarSelector);
        Delta_Weights{Layer-1} = Delta_Weights{Layer-1} + NodesActivations{Layer-1}' *
(NodesBackPropagatedErrors{Layer} .* derivative);
    end
  end

  %% Apply resilient gradient descent or/and momentum to the delta_weights
  if (enable_resilient_gradient_descent) % Handle Resilient Gradient Descent
    if (mod(Epoch,200)==0) %Reset Deltas
      for Layer = 1:nbrOfLayers
        ResilientDeltas{Layer} = learningRate*Delta_Weights{Layer};
      end
    end
    for Layer = 1:nbrOfLayers-1
      mult = Old_Delta_Weights_for_Resilient{Layer} .* Delta_Weights{Layer};
      ResilientDeltas{Layer}(mult > 0) = ResilientDeltas{Layer}(mult > 0) * learningRate_plus; % Sign
didn't change
      ResilientDeltas{Layer}(mult < 0) = ResilientDeltas{Layer}(mult < 0) * learningRate_negative; % Sign
changed
      ResilientDeltas{Layer} = max(deltas_min, ResilientDeltas{Layer});
      ResilientDeltas{Layer} = min(deltas_max, ResilientDeltas{Layer});
      Old_Delta_Weights_for_Resilient{Layer} = Delta_Weights{Layer}
      Delta_Weights{Layer} = sign(Delta_Weights{Layer}) .* ResilientDeltas{Layer};
    end
  end
  if (enable_learningRate_momentum) %Apply Momentum
    for Layer = 1:nbrOfLayers
      Delta_Weights{Layer} = learningRate*Delta_Weights{Layer} +
momentum_alpha*Old_Delta_Weights_for_Momentum{Layer};
    end
    Old_Delta_Weights_for_Momentum = Delta_Weights;
  end
```

```
if (~enable_learningRate_momentum && ~enable_resilient_gradient_descent)
    for Layer = 1:nbrOfLayers,      Delta_Weights{Layer} = learningRate * Delta_Weights{Layer};
    end
  end

  %% Backward Pass Weights Update
  for Layer = 1:nbrOfLayers-1
    Weights{Layer} = Weights{Layer} + Delta_Weights{Layer};
  end

  % Resetting Delta_Weights to Zeros
  for Layer = 1:length(Delta_Weights),       Delta_Weights{Layer} = 0 * Delta_Weights{Layer};
  end
  %% Decrease Learning Rate
  if (enable_decrease_learningRate)
    new_learningRate = learningRate - learningRate_decreaseValue;
    learningRate = max(min_learningRate, new_learningRate);
  end
%% Evaluation
  for Sample = 1:length(Samples(:,1))
    outputs = EvaluateNetwork(Samples(Sample,:), NodesActivations, Weights, unipolarBipolarSelector);
    bound = (1+unipolarBipolarSelector)/2;
    if (outputs(1) >= bound && outputs(2) < bound) %TODO: Not generic role for any number of output
nodes
      ActualClasses(Sample) = 1;
    elseif (outputs(1) < bound && outputs(2) >= bound)
      ActualClasses(Sample) = 0;
    else
      if (outputs(1) >= outputs(2)),            ActualClasses(Sample) = 1;
      else
        ActualClasses(Sample) = 0;
      end
    end
  end
```

Multilayer Perceptron Feed Forward Fully Connected Neural Network with a Sigmoid activation function$_4$

```matlab
    MSE(Epoch) = sum((ActualClasses-TargetClasses).^2)/(length(Samples(:,1)));
      if (MSE(Epoch) == 0)
         zeroRMSReached = 1;
      end

      %% Visualization
      if (zeroRMSReached || mod(Epoch,draw_each_nbrOfEpochs)==0)
         % Draw Decision Boundary
         unique_TargetClasses = unique(TargetClasses);
         training_colors = {'y.', 'b.'};
         separation_colors = {'g.', 'r.'};
         subplot(2,1,1);
         cla;
         hold on;
         title(['Decision Boundary at Epoch Number ' int2str(Epoch) '. The max number of Epochs is '
int2str(nbrOfEpochs_max) '.']);

         margin = 0.05; step = 0.05;
         xlim([min(Samples(:,2))-margin max(Samples(:,2))+margin]);
         ylim([min(Samples(:,3))-margin max(Samples(:,3))+margin]);
         for x = min(Samples(:,2))-margin : step : max(Samples(:,2))+margin
            for y = min(Samples(:,3))-margin : step : max(Samples(:,3))+margin
               outputs = EvaluateNetwork([1 x y], NodesActivations, Weights, unipolarBipolarSelector);
               bound = (1+unipolarBipolarSelector)/2;
               if (outputs(1) >= bound && outputs(2) < bound) %TODO: Not generic role for any number of
output nodes
                  plot(x, y, separation_colors{1}, 'markersize', 18);
               elseif (outputs(1) < bound && outputs(2) >= bound)
                  plot(x, y, separation_colors{2}, 'markersize', 18);
               else
                  if (outputs(1) >= outputs(2))
                     plot(x, y, separation_colors{1}, 'markersize', 18);
                  else
                     plot(x, y, separation_colors{2}, 'markersize', 18);
                  end
               end
            end
```

```matlab
         end
      end
      for i = 1:length(unique_TargetClasses)
         points = Samples(TargetClasses==unique_TargetClasses(i), 2:end);
         plot(points(:,1), points(:,2), training_colors{i}, 'markersize', 10);
      end
      axis equal;

      % Draw Mean Square Error
      subplot(2,1,2);
      MSE(MSE==-1) = [];
      plot([MSE(1:Epoch)]);
      ylim([-0.1 0.6]);
      title('Mean Square Error');
      xlabel('Epochs');
      ylabel('MSE');
      grid on;

      saveas(gcf, sprintf('Results//fig%i.png', Epoch),'jpg');
      pause(0.05);
   end
   display([int2str(Epoch) ' Epochs done out of ' int2str(nbrOfEpochs_max) ' Epochs. MSE = '
num2str(MSE(Epoch)) ' Learning Rate = ' ...
      num2str(learningRate) '.']);
   nbrOfEpochs_done = Epoch;
   if (zeroRMSReached)
      saveas(gcf, sprintf('Results//Final Result for %s.png', dataFileName),'jpg');
      break;
   end
end
display(['Mean Square Error = ' num2str(MSE(nbrOfEpochs_done)) '.']);
```

http://heraqi.blogspot.com.eg/2015/11/mlp-neural-network-with-backpropagation.html

References & Notes

Nature Publication has an impact factor 40 points (Science 35; IEEE 1.5; Opt. 3); but the most impact papers among Nature/Sci. may be delayed (~10 years esp. in physiology & physics).

1. "*Learning Machine*," Nicola Jones **V. 505**, pp146-148, **2014**;
2. "*Deep Learning*," Yann LeCun, Yoshui Bengio, Geoffrey Hinton, **V. 521**, pp. 436-440, **2015**.
3. "**Natural Intelligence Neuromorphic Engineering**," Harold Szu, Elsevier 2017, pp.1-350.
4. " **ANN, Deep Learning & Apps**," Harold Szu, Henry Chu & Simon Foo (**Gulf Mexico Spring School** April 16-19, 2017 Tallahassee FL, Elsevier Book Publisher)
5. "**Unsupervised Learning at MFE"** (single layer LCNN for one class breast cancer or not), appeared in Harold Szu, Lidan Miao, Hairong Qi, Proc. SPIE Vol. 6576, p. 657605, (2007)
6. **Multiple Layer Deep Learning** appeared in "Introduction to Computing with Neural Nets," Richard **Lipmann**, IEEE ASSP Magazine April 1987 & PDP MIT book by David **Rumelhart** , James **McCelland**; P. **Werbos** PhD Thesis.
7. **Mary Nedergaad**, **Steve Goldman (Sweden & Rochester)** "Brain Drain," Sci. Am. March 2016.
8. Nobel Prize in Medicine & Physiology has been given in 2012 to the discovery genes by Kyoto Prof. **Shinya Yamanaka**, and these **4 Yamanaka genes** can be unwind cells back to the embryonic (adult cells induced pluripotent: mice, Dolly Sheep, Homosapiens longevity)
9. **Common Sense Longevity: Sleep Tight**, **Eat Right** (calorie restriction: Luigi Fontana Wash U.)**, Deep Exercise** (e.g. Yoga,Tai Chi Quan), **Be Happy**.
10. "**Natural Intelligence Neuromorphic Engineering** (NINE)" book by Harold Szu et. al. Elsevier 2017
11. "**Deep learning ANN & Appl.**" book edited by Foo, Chu, Szu et al. from GMSS Tallahassee FL April 16-18, 2018 Elsevier 2018.